Effect of Macromolecular Crowding on the Structure and Function of a Riboswitch Aptamer in Cells and *in Vitro*

Jillian Tyrrell

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in the partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry.

Chapel Hill 2013

Approved by:

Gary J. Pielak, Ph.D.

Kevin M. Weeks, Ph.D.

Christopher J. Fecko, Ph.D.

Linda L. Spremulli, Ph.D.

Andrew L. Lee, Ph.D.

Abstract

JILLIAN TYRRELL: Effect of Macromolecular Crowding on the Structure and Function of a Riboswitch Aptamer in Cells and *in Vitro*
(Under the direction of Gary J. Pielak, Ph.D.)

There are large differences between the cellular environment and the conditions widely used to study RNA structure and function *in vitro*, but until recently no technologies allowed nucleotide-resolution analyses of RNA structure inside cells. We have now examined the structure, dynamics, and ligand-binding function of the adenine riboswitch aptamer domain in healthy, growing *Escherichia coli* cells at single-nucleotide resolution using SHAPE. We compared the in-cell RNA structure with that obtained in aqueous buffer containing 1 mM Mg^{2+}, approximately the concentration that we directly measured inside cells. The fully folded ligand-bound aptamer formed essentially the same structure in cells as in buffer. In contrast, the unbound RNA aptamer in cells was much more highly structured than the ligand-free state *in vitro*. Even high *in vitro* concentrations of Mg^{2+} did not yield the degree of structural organization observed for the free aptamer in cells. The crowded cellular environment thus stabilizes, or pre-organizes, otherwise dynamic RNA conformations significantly more than does Mg^{2+} alone, demonstrating a profound influence of the cellular environment on RNA structure.

To deduce how the structural effects observed in cells may be induced *in vitro* and how the aptamer ligand binding affinity is affected by crowding, studies in the presence of total cellular RNA, and the synthetic polymer, polyethylene glycol (PEG), were employed. These studies showed that RNA-RNA interactions and PEG had varying effects on aptamer structure. The structure of the aptamer in the presence of both cellular RNA and large PEGs agreed better with the structure observed in cells. However, specific characteristics observed for the aptamer in cells were not observed *in vitro* under either condition. Crowding the aptamer with PEG had little effect on ligand binding affinity, even when large structural effects were observed. These results suggest that the complex cellular environment is difficult to mimic *in vitro*, and to do so, may require a combination of crowding agents having different chemical properties. Finally, the observed minimal effects on ligand binding affinity for the aptamer in the presence of PEG, suggests that while RNA structure is affected by crowding in cells, RNA function may not be significantly affected.

“Go then, there are other worlds than these.”

— Stephen King, *The Gunslinger*

Acknowledgements

Special thanks to Gary J. Pielak for being an outstanding mentor and advisor. Gary has always been available and supportive, and I am grateful to have been part of his lab through this process. Thanks to my fellow Pielak group members for their support and friendship.

Many thanks to Kevin M. Weeks, our collaborator on this project, who has always been very helpful and a great mentor to me as well. Also, I would like to thank Kevin and the Weeks lab members for allowing me to be a pseudo Weeks lab member, and including me in their group meetings and other lab functions.

It has been a long road pursuing my goals and I would like to thank my son, Bradly, and my husband, Josh, for their love and support. My son has grown up watching me work very hard for my achievements. He was 5 years old when I started on my undergraduate degree, and now he is 16, a wonderful young man, and ready to begin his own journey soon. I hope he has learned the value of education, but more importantly, that if he works hard, he can achieve anything he wants.

Table of Contents

List of Figures

List of Abbreviations and Symbols

°C	degree Celcius
µL	microliter
µM	micromolar
1M7	1-methyl-7-nitroisatoic anhydride
2AP	2-aminopurine
2PEF	two-photon excitation fluorescence
3PEF	three-photon excitation fluorescence
5-FAM	5-carboxyfluorescein
6-HEX	6 - carboxy - 2',4,4',5',7,7' - hexachlorofluorescein
A	adenine
AM	acetoxymethyl
BzCN	benzoyl cyanide
cDNA	single-stranded, complementary deoxyribonucleic acid, product of reverse transcription
C	cytosine
DAP	diaminopurine
ddATP	dideoxyadenosine triphosphate
dNTP	deoxynucleotide triphosphate
DMAP	dimethylaminopurine
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid

FPLC	fast performance liquid chromatography
g	gram
x g	times gravity
G	guanine
h	hour
HEPES	N-2-hydroxyethylpiperazine-N'-ethanesulfonic acid
In vitro	Latin for experiments performed in a test tube
In vivo	Latin for experiments performed in living organisms
IPTG	isopropyl β-D-1-thiogalactopyranoside
J1-2, J2-3, J3-1	junctions between helices in the adenine riboswitch aptamer
kDa	kilodalton
K_D	equilibrium dissociation constant
L	liter
L2, L3	loop regions 1 or 2 in the adenine riboswitch aptamer
LNA	locked nucleic acid
LB	Luria broth
lys	lysine
M	molar
mg	milligram
Mg^{2+}	magnesium ion
$MgCl_2$	magnesium chloride
min	minute
mM	millimolar

MPEF	multi-photon excitation fluorescence
mRNA	messenger RNA
NaCl	sodium chloride
nm	nanometers
nM	nanomolar
NMR	nuclear magnetic resonance spectroscopy
OD_{600}	optical density at 600 nanometers
P(1-3)	helix 1, 2, or 3 in the adenine riboswitch aptamer
PAGE	polyacrylamide gel electrophoresis
PEG	polyethylene glycol
PVP	polyvinylpyrrolidone
R	Pearson's Correlation Coefficient
RNA	ribonucleic acid
RT	reverse transcription
s	second
SDS	sodium dodecyl sulfate
SHAPE	Selective 2'-Hydroxyl Acylation analyzed by Primer Extension
tRNA	transfer ribonucleic acid
U	uracil
vol	volume
w/v	weight/volume

Chapter 1: Introduction

1.1 Significance of Macromolecular Crowding

Many cellular RNA functions require formation of sophisticated secondary and tertiary structures. Extensive work, largely performed under simplified conditions, has shown that higher-order RNA structure is exquisitely sensitive to the surrounding solution environment. Most *in vitro* explorations of RNA molecules are carried out in dilute buffer with minimal components. However, the environment in cells, where RNA molecules function, is extraordinarily complex and contains concentrations of macromolecules upwards of 300 g/L.[1] Crowding effects can arise from two phenomena: hard-core repulsions and non-specific chemical interactions. Hard-core repulsions arise from volume exclusion or charge repulsion and result in reduced conformational space, which favors compaction.[2,3] In contrast, non-specific chemical interactions, such as attractive charge-charge interactions, can destabilize compact, folded states. In cells, it is likely that both forces affect the structure and function of RNA.

While the effects of crowding on proteins have been studied extensively,[4,5] effects on RNA are only recently being investigated.[6,7] It is plausible that macromolecular crowding may have a substantial effect on RNA structure and function. First, unlike most proteins, RNA carries a high negative charge resulting from the phosphate backbone. RNA structure may therefore be affected significantly

by either repulsive or attractive charge-charge interactions. Additionally, many functional RNA's have highly dynamic regions that sample multiple conformations;[8] hard-core repulsions will favor the more compact members of this ensemble.

Since it is difficult to mimic the cellular environment, studies on the effects of crowding on RNA have been limited to synthetic crowding agents, such as polyethylene glycol (PEG), and trimethylamine oxide (TMAO).[6,7,9,10] Most of these studies focus only on global effects on RNA structure, such as the radius of gyration, and there are few studies that give insight into crowding effects on function. To fully understand RNA molecules and their structure and function relationships, functional assays, and studies under physiological conditions at nucleotide resolution, are required.

1.2 Riboswitches exemplify the relationship between RNA structure and function

Riboswitches are regulatory genetic elements located in the 5'-untranslated region of several prokaryotic mRNAs.[11] Typically, riboswitches comprise an aptamer domain and an expression platform. The aptamer domain secondary structure is highly conserved and undergoes a conformational change in response to specific ligand binding. This conformational change results in gene expression control by either formation or disruption of transcription terminators or hairpins that contain transcription or translation initiation signals in the expression platform. Thus, riboswitch function depends on its structure.

1.2.1 *add* A riboswitch

Purine riboswitches are highly conserved between organisms and have been well characterized.[12-14] The purine class includes riboswitches that recognize and bind guanine or adenine. Distinctly, discrimination between guanine and adenine for each riboswitch depends on a single-nucleotide, located in the binding pocket of the aptamer domain. One member of the purine riboswitch class, the *add* adenine riboswitch, acts as an 'on' switch for the expression of adenine deaminase. Specifically, when cellular adenine concentrations reach a threshold, adenine binding to the aptamer causes a conformational change that turns 'on' transcription of the gene encoding adenine deaminase by exposing Shine-Delgarno sequence (Figure 1.1).

The *add* A riboswitch aptamer binds adenine and analogs 2-aminopurine and 2, 6-diaminopurine with high affinity (K_D~ 60 nM, 117 nM, and 2 nM, respectively).[13,15] The ligand-bound aptamer is a highly organized structure where long-range tertiary interactions between loops 2 and 3 (L2 and L3) are intact and the ligand is completely encapsulated and involved in extensive tertiary interactions within the binding pocket (Figure 1.2).[12,13,16] In contrast, the ligand-free aptamer is dynamic, having a disordered binding pocket and only transient interactions between L2 and L3.[16]

1.3 SHAPE Chemistry

Few methods allow exploration of intracellular RNA structure. The most commonly used approach employs the chemical probing reagent, dimethyl sulfate

(DMS). Studies using DMS have yielded significant insights to the intracellular structure of several RNAs.[17,18] However, DMS reacts with only a few functional groups on RNA, primarily adenosine (N1) and, to a lesser extent, cytosine (N3). *In vivo* DMS probing studies are therefore best coupled with information from other techniques.[18]

An alternative chemical probing method, selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) employs electrophilic reagents that react at the 2'-hydroxyl group of conformationally flexible nucleotides to form 2'-*O*-adducts (Figure 1.3) and yields quantitative information on the degree to which a nucleotide is constrained by base pairing, or other interactions. SHAPE has been widely used to develop secondary structure models and detect complex conformational changes of RNA *in vitro*.[19-23] Here, we explore the use of SHAPE to probe RNA structure in healthy, growing *E. coli* cells and compare our results directly with *in vitro* SHAPE experiments

The 1-methyl-7-nitroisatoic anhydride (1M7) SHAPE reagent provides a straightforward approach for probing RNA structure in living cells on a minute timescale. The results presented here are not compatible with a recent report suggesting that 1M7 does not perform well in cells and that slowly reacting reagents are required for in-cell probing. [24] SHAPE reagents that react rapidly with RNA, as is the case with 1M7,[12] versus the recently described reagents that react slowly,[24] likely probe different features of RNA structure. In general, fast reacting reagents do a better job of quantifying intrinsic RNA structure, while slow reagents are sensitive to and yield information related to slow dynamic features of an RNA.[25] The intrinsic

reactivity of slow reacting reagents appears to highly sensitive to the specific *in vitro* ion concentrations, which will complicate understanding differences between in-cell versus *in vitro* measurements. A fast reacting reagent is thus especially important when examining the effects of the cellular environment on RNA structure.

In this method, I add 1M7 to *Escherichia coli* cells expressing the *add* A aptamer or to purified aptamer, *in vitro*. We locate positions of 2'-hydroxyl modifications by first using sequence-specific, fluorescent primers to reverse transcribe the modified RNA. The enzyme used, reverse transcriptase, will stop at modified nucleotides. Thus, the reverse transcription (RT) reaction yields a pool of cDNA of different lengths corresponding to the location of modified nucleotides. Each experiment includes an RT reaction with unmodified RNA to determine background stops. Second, we apply capillary electrophoresis to quantify cDNA products. The electropherograms are aligned with results from a dideoxynucleotide sequencing reaction to determine peak positions. Peak intensities are integrated and corrected for background and signal decay to yield a SHAPE profile.

My results show that 1M7 readily traverses the complex multi-layered cell wall structure of *E. coli* to yield nucleotide-resolution structural and ligand-binding information for the *add* A riboswitch aptamer domain.

1.4 Overview: effects of crowding on RNA structure and function

In this dissertation, I focus on determining the consequences of macromolecular crowding in cells and *in vitro* on the structure of the *add* A riboswitch aptamer domain. In Chapter 2, in-cell SHAPE is used to probe both the

ligand-free and ligand-bound aptamer structure in healthy, growing *E. coli*. I apply aptamer ligand binding assays and SHAPE *in vitro* in the presence of synthetic crowding agents in Chapter 3, to help tease apart the consequences of hard-core repulsions and non-specific interactions on RNA structure and function. The results give insight into the consequences of the crowded cellular environment on RNA structure and function.

1.4.1 Effects of the cellular environment on *add* A aptamer structure

RNA forms complex secondary and tertiary structures that are essential for cellular function. I set out to determine if the crowded cellular environment had an effect on the structure of the free and ligand-bound aptamer. The ligand-bound aptamer structure, observed crystallographically, is compact and highly organized[13] and therefore not expected to be perturbed by macromolecular crowding forces. In contrast, the free aptamer in the presence of physiologically relevant Mg^{2+} concentration has a disordered binding pocket, and is lacking stable loop-loop interactions.[16,26] If hard-core repulsions dominate in cells, a more ordered, and compact state will be observed for the aptamer in cells compared to in buffer. Alternatively, if non-specific attractive interactions dominate, I will observe a structure more disordered than that observed in buffer with physiological Mg^{2+}.

My results for the ligand-bound aptamer were as expected: the cellular environment had little effect on the structure. In contrast, SHAPE profiles for the free aptamer show that the cellular environment indeed has a significant effect on the adenine aptamer structure. The impact of the cellular environment on individual motifs differs, indicating that a combination of attractive chemical interactions and

hard-core repulsions play a role in the structural effects. However, the overall structure is more compact and organized in cells. Thus, hard-core repulsions are the dominant force responsible for the differences in aptamer structure in cells compared to in buffer. These results underscore the importance of examining the consequences of the complex cellular environment on RNA structure and function in cells.

1.4.2 Effects of crowding on *add* A aptamer ligand binding *in vitro*

Riboswitches harness structural changes induced by ligand binding to control genes. Crowding effects on structure would alter the equilibrium between the free and bound states, thereby affecting the equilibrium binding constant (K_D). Currently, there are no methods to measure K_D directly in cells. For these studies, I employed synthetic polymers to measure the consequences of crowding on K_D for the *add* A riboswitch aptamer:2AP complex.

These studies show that while crowding has very little effect on K_D for the aptamer:2AP complex, crowding agents of different size, geometry, and chemical composition have varying effects. Understanding the nature of the interactions between crowder and aptamer shed light on how hard-core repulsions and attractive chemical interactions may affect the thermodynamics for RNA ligand binding and suggest that the in-cell K_D may not agree with that measured *in vitro*.

1.5 Future Directions

The goal of this work is to understand the structure and behavior of RNA in its natural environment, the cytoplasm. From my results, we now know the

conformation of the *add* A riboswitch aptamer in cells differs from that observed *in vitro*. A significant contribution to understanding RNA biophysics would be to establish conditions *in vitro* that can induce the conformation observed in cells. Such knowledge would not only further our understanding of the specific contributions the cytoplasm has on RNA structure, but also allow RNA to be studied under 'physiological like' conditions *in vitro*. This ability would permit studies of much larger RNA's that cannot effectively be studied in cells.

Additionally, this work provides insight into how hard-core repulsions and chemical interactions affect the ligand-binding thermodynamics of the *add* A riboswitch aptamer. With this knowledge, we could deduce how RNA ligand binding thermodynamics is affected in cells. Finally, the ultimate goal would be to develop a technique to measure ligand binding affinity directly in cells.

1.6 Figures

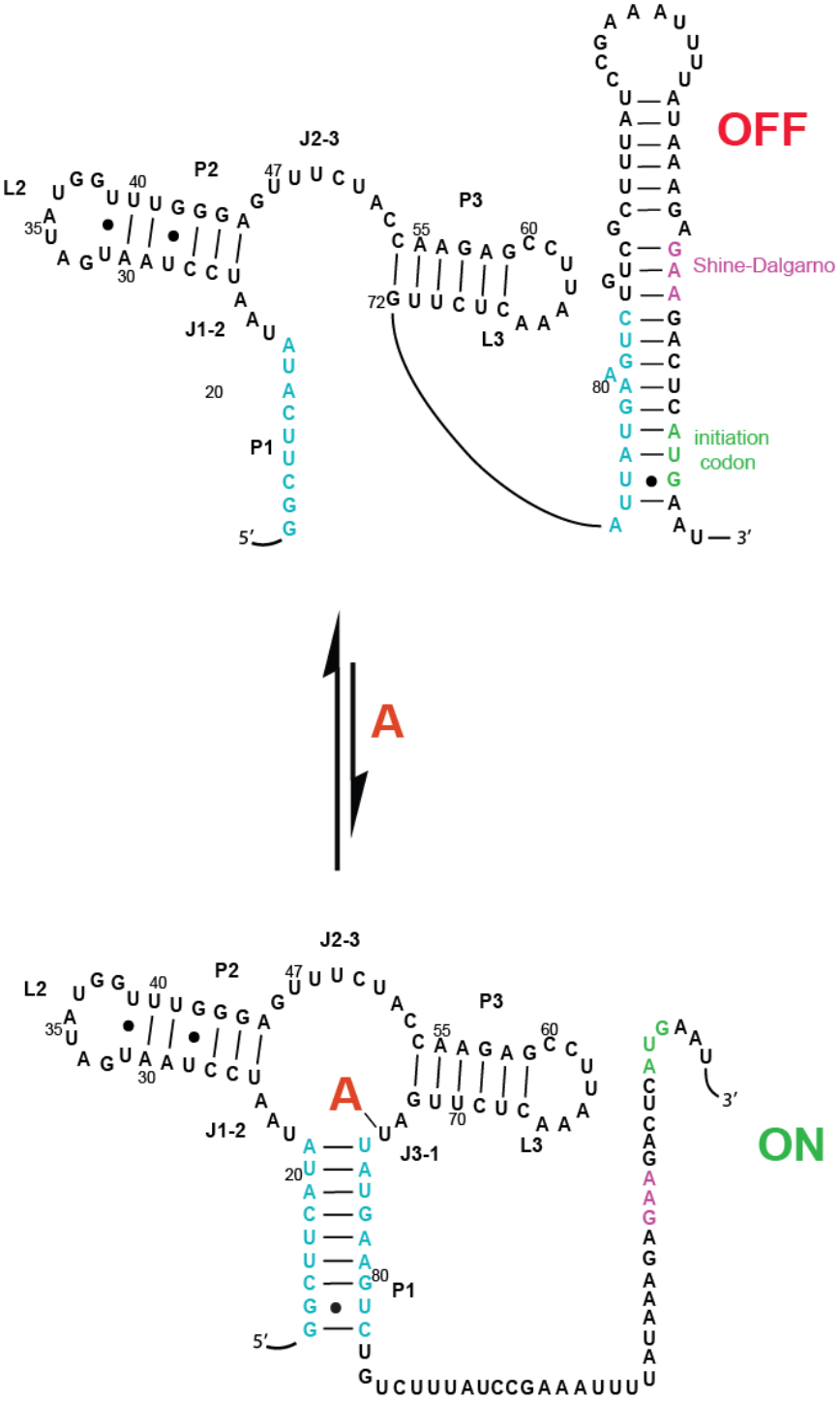

Figure 1.1 Mechanism for Gene Regulation for the *add* A Riboswitch. Adenine binding induces a conformational change, which exposes the Shine-Dalgarno sequence and initiation codon, turning on translation of adenine deaminase.

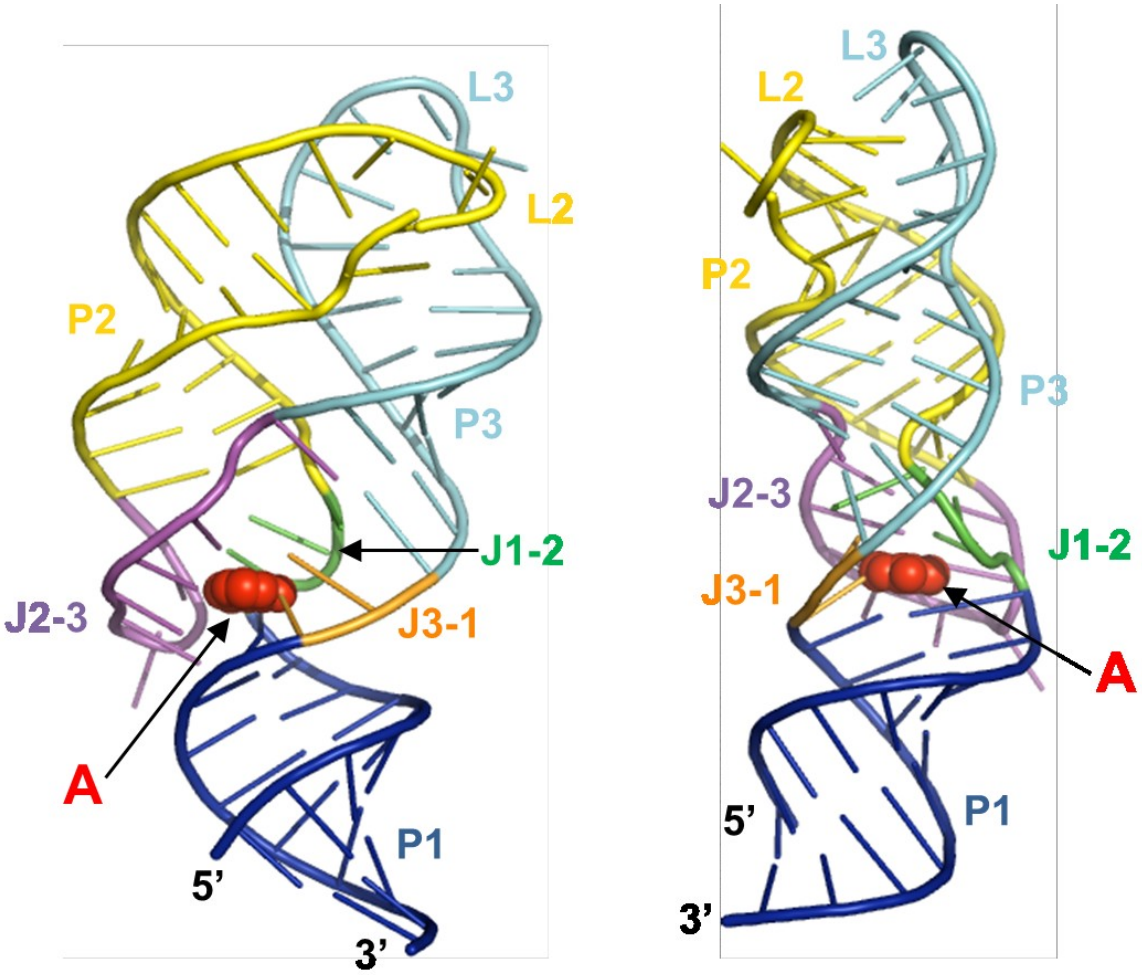

Figure 1.2 Cartoon Representations of the *add* A Riboswitch Aptamer-Adenine complex (PDB Y126).[13]

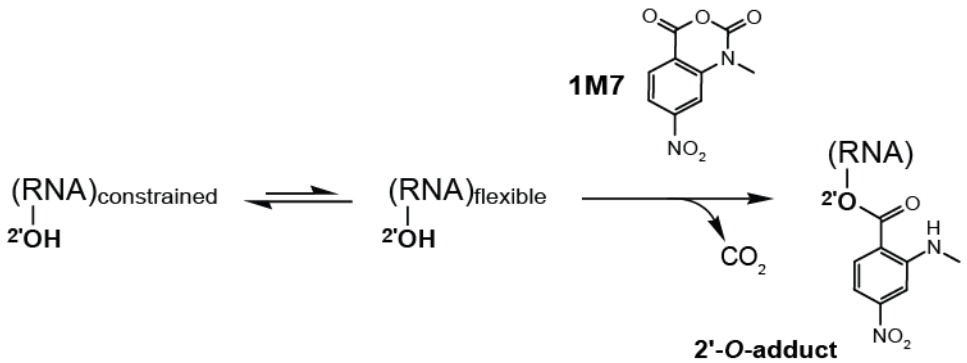

Figure 1.3 SHAPE Mechanism.Structure-selective 2'-hydroxyl acylation using 1M7. 2'-O-adducts are detected using primer extension and capillary electrophoresis.

Chapter 2- Pre-Organization of RNA Structure in Living Cells

The material in this chapter is from:

Tyrrell J, McGinnis JL, Weeks KM, Pielak GJ, Effects of the Cellular Environment on Adenine Riboswitch RNA Structure, 2013.

(J.T., J.L.M., K.M.W, and G.J.P. designed research; J.T. and J.L.M performed research; J.T., J.L.M., K.M.W, and G.J.P. analyzed data; J.T., K.M.W, and G.J.P. wrote the paper.)

2.1 Introduction

The cellular environment is vastly different from that employed in most *in vitro* explorations of biological macromolecules. In bacterial cells, macromolecules reach concentrations of 300 g/L and occupy up to 30% of the total volume.[1] This crowding leads to two types of interactions between macromolecules: hard-core repulsions and chemical interactions. Hard-core repulsions reflect the impenetrable nature of atoms and reduce conformational space, ultimately favoring more compact states.[2,3] Chemical interactions can be attractive or repulsive. Repulsions arise from interactions between like charges, reinforcing the hard-core effect. Attractive interactions, including hydrogen-bonding and juxtaposition of complementary charged species, are expected to destabilize the folded state because they are favored by exposing the local structures that mediate these interactions.[27] Biological molecules are likely impacted by both phenomena in cells; however, the extent of these contributions is poorly understood because the complexity of the cellular

environment is difficult to mimic *in vitro.* Studies of biological molecules in realistic cellular environments are essential for fully understanding the impact of complex cellular contributions to macromolecular structure and function.[28]

RNA molecules are responsible for diverse cellular functions, many governed by precise features of their three-dimensional structure. *In vitro* studies show that crowding induced by synthetic polymers has dramatic consequences for RNA folding.[6,7,10] In addition, the negatively charged backbone of RNA makes folding exquisitely sensitive to cation concentrations.[29] Despite these considerations, most *in vitro* explorations of RNA structure and function employ purified RNAs and simple buffered solutions at Mg^{2+} concentrations that are significantly higher than occur physiologically.

To probe the effect of the cellular environment on RNA structure, we expressed a riboswitch aptamer domain in *E. coli* cells using a tRNA scaffold.[30,31] We compared the reactivity of virtually all nucleotides in the aptamer using SHAPE chemistry in healthy, growing cells and in buffer. SHAPE yields quantitative information on the degree to which a nucleotide is constrained by base pairing or other interactions and has been widely used to develop secondary structure models and detect complex conformational changes in RNA *in vitro*.[19,20,32,33] Application of SHAPE to living cells promises to significantly improve our ability to understand how the cellular environment affects RNA structure at the level of individual nucleotides.

Riboswitch RNAs regulate gene expression by undergoing ligand-induced conformational changes that ultimately enhance or inhibit expression of a linked gene.[11] The *add* A riboswitch has been well characterized *in vitro*. In the absence of

ligand, the aptamer domain has a disordered binding pocket; upon ligand binding, stable tertiary interactions form (Figure 2.1A).[12,16,34] Using a fast-acting SHAPE reagent,[35] we found that the ligand-free aptamer has a more compact and more highly ordered structure in the cellular environment than does the purified RNA in buffer at physiological Mg^{2+} concentrations. Moreover, the conformation that predominates in cells could not be achieved *in vitro*, even in buffers containing 30 times the concentration of Mg^{2+} directly measured in cells. Our results indicate that macromolecular crowding stabilizes, or pre-organizes, tertiary interactions and emphasize the importance of studying RNA structure in cells.

2.2 Results

2.2.1 Intracellular Mg^{2+} concentration.

RNA tertiary structure is highly dependent on the Mg^{2+} concentration.[26,29] To ensure accurate comparison of the effect of this ion on RNA structure, we measured the free Mg^{2+} concentration in *E. coli* cells using a ratiometric Mg^{2+}-selective fluorophore.[36] Cells were grown to late-log phase and intracellular Mg^{2+} concentration was measured as a function of external Mg^{2+} concentration in physiological buffer [10 mM HEPES-KOH (pH 7.4), 200 mM potassium acetate, 5 mM NaCl]. At an external concentration corresponding to that in standard LB medium [0.3 mM, calculated from the Mg^{2+} content of tryptone and yeast extract[37]] the intracellular Mg^{2+} concentration was 0.8 ± 0.2 mM (Fig. 2.2, arrow). The intracellular Mg^{2+} concentration varied only slightly when the extracellular ion concentration was increased. The intracellular free Mg^{2+} concentration did not rise

above 1 mM until the extracellular concentration was 5 mM (1.3 ± 0.3 mM) and reached 2.2 ± 0.5 mM at an external concentration of 10 mM. The in-cell Mg^{2+} concentration measured here for *E. coli* (0.8 ± 0.2 mM) is consistent with measurements obtained for other cell types.[38-41]

2.2.2 In-cell structure of the adenine aptamer domain in *E. coli*.

To study the aptamer domain in *E. coli* cells, we expressed the RNA as a chimera in which the aptamer domain was inserted into the anticodon loop of $tRNA^{lys}$ (Figures 2.1A). These chimeric RNAs fold stably in cells, are processed like normal tRNAs, and yield an intact and monodisperse product.[30,31] The RNA containing the aptamer was processed into a product of approximately 143 nucleotides, indicating that the tRNA domain folds correctly in cells (Figure 2.1B).

In-cell SHAPE probing was achieved by adding 1-methyl-7-nitroisatoic anhydride (1M7) in dimethyl sulfoxide (DMSO) to late log-phase cells, under conditions that are fully compatible with normal cell growth and viability. Measurements of 1M7 hydrolysis in LB at 37 °C and pH 7.0, corresponding to culture media at late log phase, showed that 1M7 has a half-life of 0.2 min (Figure 2.3); thus, SHAPE probing yields a profile of cellular RNA structures over an approximately 2-min interval. We have consistently found that 1M7 is straightforward to use *in vivo*, including in bacterial cells. Our experience is not compatible with a recent report suggesting that 1M7 performs poorly in cells and that slowly reacting SHAPE reagents are required for in cell probing.[24] In general, prior work indicates that slowly reacting reagents are likely to be poor choices for *in vivo* studies. Reagents like 1M7 that react rapidly with RNA and reagents that react slowly probe

different structural features in RNA. In general, fast reacting reagents provide quantitative measurements of intrinsic RNA structure, whereas slow reagents are sensitive to slow dynamic features of an RNA.[25,42,43] A critical consequence of this difference is that slow reacting reagents are highly sensitive to *in vitro* ion concentrations,[35] whereas fast reacting reagents are not. Thus, the use of a fast reacting reagent makes possible direct comparison of in-cell and *in vitro* measurements without need for corrections based on ion sensitivity.

After modification, total cellular RNA was isolated, and sites of aptamer modification were determined after primer extension by capillary electrophoresis.[32] SHAPE probing with 1M7 yielded electropherograms comparable or better in quality to those obtained from *in vitro* experiments (Figure 2.4). Little or no degradation of the adenine riboswitch RNA was observed in the cells or in RNA treated *in vitro* with DMSO. Thus, signals were due to reactivity of the RNA with 1M7, and the 1M7 reagent readily diffuses across the double walled membrane of the Gram-negative *E. coli* bacterium.

We first evaluated the SHAPE reactivity profile in cells with no added ligand. In-cell In the SHAPE reactivity profile of the free aptamer (Figure 2.5, black trace), nucleotides expected to form helices P1, P2, and P3 (Figure 2.1A) had low reactivity, consistent with stable base pairing. Most nucleotides that form the ligand binding pocket (J1-2, J2-3, J3-1) were moderately (≥0.4) or highly (>0.8) reactive as were nucleotides in the loop regions L2 and L3. Nucleotide U48 was hyper-reactive. The in-cell SHAPE reactivity profile for the free aptamer suggests the three major RNA

helices are formed and that the ligand-binding pocket is relatively flexible and partially disordered.

2.2.3 Ligand-free aptamer is more highly organized in cells than in buffer.

We observed large differences when we compared the in-cell riboswitch unliganded aptamer structure to that in buffered solution at 1 mM Mg^{2+} *in vitro* (Figure 2.5, compare orange and black). Most nucleotides in the adenine binding pocket (regions J2-3 and J3-1) were less reactive in cells than *in vitro*, consistent with formation of stacking and base pairing interactions in these regions in the cell and the lack of these interactions in buffer. Interestingly, the J1-2 region was more reactive in cells than in buffered solution. This higher reactivity may arise from increased solvent accessibility due to tertiary collapse and ordering of J2-3 and J3-1 and/or to attractive chemical interactions with other cellular components. Finally, nucleotide U48 was hyper-reactive in cells but showed only high reactivity in buffer. In the ligand-bound RNA in the crystal structure, U48 is flipped into the solvent [13] and is constrained in one of the (relatively rare) conformations that activates the ribose 2'-hydroxyl group for SHAPE chemistry (Figure 2.6).[44] Nucleotides G37 and G38 in L2 were also significantly more reactive *in vitro* than in cells, suggesting that tertiary interactions involving loops L2 and L3[13,26] are largely absent in buffered solution but are at least partially formed in cells. Overall, these data show that the aptamer is more structured and has a more highly organized binding pocket in cells than it does in buffer at a physiologically relevant Mg^{2+} concentration.

2.2.4 Ligand binding to RNA in cells.

The *add* A riboswitch recognizes 2-aminopurine (2AP) and adenine with similar affinities [K_d ~ 117 nM and 60 nM, respectively,[15,45,46]], and both ligands induce large conformational changes in the riboswitch RNA.[12,15,46] 2AP, however, is not found in detectable quantities in *E. coli,* and the expression level of our RNA construct is ~100 times (see Methods and materials) the intracellular concentration of adenine (~10^{-6} μM).[47] Thus, using 2AP, it is possible to examine the in-cell consequences of ligand binding without interference by endogenous cellular metabolites.

When 2AP was added to growth medium, we observed a substantial change in SHAPE reactivities (Figure 2.7A). The changes were consistent with ligand-mediated conformational changes in the RNA and formation of the intricate network of base-pairing (Figure 2.1A) and stacking interactions observed by crystallography[13] and NMR.[16] Specifically, reactivities of nucleotides in the binding pocket decreased significantly (Figure 2.7A, regions J1-2 and J3-1). The J1-2 region, which is highly reactive in the free state, became almost completely unreactive. The reactivity of U48 increased two fold, consistent with the ligand-bound RNA crystallographic structure in which this position is constrained in a conformation that facilitates SHAPE reactivity (Figure 2.6).[13,44] Reactivity decreased at G38 in L2, which pairs with C61 in L3. Reactivity also increased at U62 and U63, consistent with the crystal structure, which shows that these nucleotides are unpaired and likely dynamic in the ligand-bound state.[13]

With the exception of nucleotides in the J2-3 region, which were more reactive in cells, the SHAPE profiles of ligand-bound RNA in cells and in buffer are very similar (compare Figure 2.7A and 2.8). Thus, the ligand-bound structure is not as strongly affected by the cellular environment as the ligand-free aptamer. Two additional ligand controls were performed to assess the specificity of the conformational changes induced by 2AP. The nucleotide 2,6-diaminopurine binds the aptamer with high affinity ($K_D \approx 2$ nM)[15] and, as expected, induced large changes in the in-cell aptamer SHAPE reactivity profile, comparable to those induced by 2AP (Figure 2.9). *N*6, *N*6-Dimethyl-adenine does not bind the aptamer ($K_D > 300$ μM)[12] and the in-cell SHAPE profile obtained from cells grown in the presence of this molecule was similar to that of the free aptamer (Figure 2.9). These results emphasize that conformational changes induced by 2AP reflect specific binding in the ligand binding pocket of the riboswitch.

2.2.5 Pre-organization of the free aptamer in cells is not due to binding by pre-existing ligand.

To confirm that differences observed between in-cell and *in vitro* SHAPE reactivity profiles of the free aptamer were not due to endogenous adenine binding, we investigated a non-binding mutant. Nucleotide U74 forms a canonical base pair with the adenine ligand and is critical for ligand affinity and specificity.[12,46] Therefore, the U74G mutation should disrupt binding to adenine, 2AP, and related molecules. The SHAPE profiles of the U74G mutant in buffer at 1 mM Mg^{2+} both in the absence and presence of 2AP were essentially identical (Figure 2.10). Thus, the ligand-free U74G mutant does not bind 2AP *in vitro*.

In-cell analysis of the U74G mutant in the absence and presence of added 2AP in the cell media also yielded nearly identical SHAPE profiles (Figure 2.7B), indicating that the mutant does not bind ligand in cells. Critically, the in-cell pattern and magnitude of SHAPE reactivities for the binding pocket and loop regions were also nearly identical to those of the native sequence RNA. Thus, the non-binding mutant possesses the same partially organized state in cells observed for the native sequence in the absence of ligand. Therefore, the structure of the unliganded native-sequence aptamer in cells does not arise from endogenous adenine ligand binding, but instead indicates a definitive role of the cellular environment in pre-organizing the RNA.

2.2.6 High Mg^{2+} concentrations do not induce the pre-organization observed in cells.

In vitro experiments performed at Mg^{2+} concentrations comparable to those measured in *E. coli* (Figure 2.2) indicated that the adenine riboswitch aptamer domain was notably less structured than in cells than *in vitro*. We therefore attempted to induce the highly ordered in-cell conformation by increasing the Mg^{2+} concentration *in vitro*. We probed the aptamer structure as a function of Mg^{2+} concentration and calculated the Pearson correlation coefficient (*R*) between in-cell and *in vitro* SHAPE reactivities (Figure 2.11). The correlation between the in-cell reactivities and those measured in buffer at 1 mM Mg^{2+} was low (0.55, inset in Figure 2.11), consistent with large differences in RNA structure under these two conditions. Increasing the Mg^{2+} concentration to 2 mM, increased the correlation to 0.69, and increasing the ion concentration to 30 mM yielded a correlation of 0.78 (Figure 2.11). In sum, the conformation of the free aptamer RNA in cells is uniquely

stabilized by the cellular environment, and this contribution cannot be recapitulated in simple buffers by increasing the concentration of Mg^{2+}.

2.3 Discussion

The adenine riboswitch allows bacterial cells to control gene expression in the adenine biosynthesis pathway in response to purine levels in the environment. Like the functions of many non-coding RNAs, riboswitch function is mediated by its higher-order structure. The crowded and unique ion environment in cells is likely to have a large impact on RNA structure-function relationships, but these effects remain poorly understood. Using in-cell SHAPE, performed in 2-min structural snapshots (Figure 2.3), we found that the cellular environment induces a collapsed conformation in the unbound adenine riboswitch aptamer RNA that is different from the ligand-bound structure and from the structure *in vitro*, even when the buffer included high Mg^{2+} concentrations.

We visualized the overall effects of the cellular environment on the level of structural organization of the RNA aptamer using a scale based on differences in SHAPE reactivities. Values between zero and one provide a metric of structural organization in individual RNA motifs relative to the fully folded and highly organized, ligand-bound state. SHAPE reactivities for nucleotides involved in the long-range loop-loop interaction and for the three motifs that comprise the ligand-binding pocket are diagnostic of the overall structure (Figure 2.12, right). Binding by the 2-aminopurine ligand induced a highly structured RNA state both in buffer and in cells as indicated by structural organization values near 1 for both the tertiary and ligand-binding motifs (Figure 2.12, solid symbols). Region J2-3 was slightly less organized

in cells than *in vitro*; whereas, tertiary interactions in L2 and in binding pocket regions J1-2 and J3-1 were slightly more constrained in cells. The in-cell ligand-bound structure thus appears to be slightly more stable than that formed *in vitro*.

In cells, the ligand-free aptamer was much more highly structured than was the free RNA in buffered solution *in vitro* (Figure 2.12, compare open stars and open rectangles). In cells, with the exception of nucleotides in J1-2, the free RNA was almost as structured as the 2AP-bound state (structural organization values were 0.8). When the Mg^{2+} concentration was increased to 10 mM for *in vitro* experiments (Figure 2.12, open circles), long-range tertiary contacts involving the L2 loop were stabilized, but the effect on the ligand binding pocket was small. The non-binding U74G mutant (Figure 2.7B) had the same highly ordered structure as the ligand-free native sequence aptamer, emphasizing that in-cell stability does not arise from endogenous ligand binding. The cellular environment thus significantly stabilizes tertiary interactions and the binding pocket in the ligand-free aptamer in a way that is not replicated by high Mg^{2+} concentrations *in vitro*.

In summary, we have used SHAPE chemistry with the well-understood 1M7 reagent[35,44,48] to study the structure of the adenine riboswitch aptamer RNA in living cells at single-nucleotide resolution. We found that the cellular environment 'pre-organizes' the riboswitch RNA for ligand binding. The high level of organization specifically observed in the ligand-binding pocket (Figure 2.12) for the free aptamer in cells suggests that the free energy difference between the ligand-free and ligand-bound conformations is less inside the cell than in buffered solution. In principle, RNA structure is influenced by both hard-core repulsions and repulsive and

attractive chemical interactions. Our observation that the aptamer RNA is highly structured in cells suggests that hard-core and charge-charge repulsions dominate. Many RNAs show low levels of function under physiological Mg^{2+} concentrations *in vitro* but are more functional at high ion concentrations. As we observe here, however, the structure stabilized *in vitro* by high Mg^{2+} concentrations may not fully recapitulate the functional structure in cells.

2.4 Materials and Methods

2.4.1 Concentration of Mg^{2+} in *E. coli*.

Free Mg^{2+} was measured using the intracellular chelator, mag-fura-2, acetoxymethyl (AM) (Figure 2.13). The free Mg^{2+} concentration is obtained from the ratio of the fluorescence intensities of the free and complexed dyes at 510 nm measured using excitation 380 nm and 340 nm, respectively. Methods were adapted from those described.[41]

E. coli BL21(DE3) cells were grown in LB media at 37 °C, with shaking, until the optical density at 600 nm (OD_{600}) reached ~1. Aliquots (12 mL) were centrifuged at room temperature for 15 min at 1500 × *g*, washed once, and resuspended in 1.5 mL Mg^{2+}-free buffer [10 mM HEPES-KOH (pH 7.4), 200 mM potassium acetate, 5 mM NaCl]. Cells were incubated with shaking at 37 °C for 10 min, then 500 µl of dye mixture containing 20 µM mag-fura-2 AM and 60 µM Pluronic F-127 (Molecular Probes). Dyes were prepared in Mg^{2+}-free buffer from mag-fura-2 AM (5 mM, in anhydrous DMSO) and Pluronic F-127 [20% (w/v) in DMSO] stocks. After dye addition, cells were incubated for 70 min at 37 °C with shaking. Cells were then

washed, resuspended in 2 mL of Mg^{2+}-free buffer, and incubated for 30 min to ensure hydrolysis of the intracellular dye. Cells were then washed twice and resuspended in Mg^{2+}-free buffer to yield an OD_{600} ~0.5 (~2 × 10^8 cells/mL). Cell viability was determined by plating serial dilutions of cell suspensions on LB-agar plates containing 1 mg/mL ampicillin.

The fluorescence of cell suspensions (2.4 mL in a stirred 3 mL cuvette) was measured at 37 °C using a Varian Cary Eclipse fluorimeter. Excitation at 340 nm and 380 nm was alternated at 1 s intervals, and emission was measured 510 nm. To quantify Mg^{2+}-dependent changes in fluorescence, Mg^{2+} was added to cells at 100 s intervals (Figure 2.13A). The intensities and 340 nm/380 nm ratios were used to calculate the internal Mg^{2+} concentration.[36]

$$[Mg^{2+}] = K_D\,(F_0/F_S)\,(R_{Mg}{}^{2+} - R_{min})/(R_{max} - R_{Mg}{}^{2+}) \qquad (1)$$

where K_D is the dissociation constant of the mag-fura-2/Mg^{2+} complex (2.5 mM, see below), F_0/F_S is the ratio of intensities with 380 nm excitation for 0 mM added Mg^{2+} and saturating Mg^{2+}, $R_{Mg}{}^{2+}$ is the ratio of 510 nm emission from excitation at 340 nm and 380 nm at a given external Mg^{2+} concentration, and R_{min} and R_{max} are the minimum and maximum ratios determined at 0 mM and saturating Mg^{2+}, respectively. R_{min} and R_{max} were obtained at the end of each experiment by adding Mg^{2+} to a final concentration of 30 mM, then adding SDS [to 0.1% (w/v) final] to lyse the cells. Lysis causes mag-fura-2 release from cells and saturatation with Mg^{2+} (Figure 2.13B). This step was followed by addition of EDTA to a final concentration

of 60 mM to chelate Mg^{2+} and achieve a free Mg^{2+} measurement.

The K_D was determined by recording Mg^{2+}-dependent fluorescence changes for hydrolyzed mag-fura-2 (0.4 μM) in buffer and fitting the 510 nm intensities from 340 nm excitation to a one-site binding curve (Figure 2.14).

Emission from cells in Mg^{2+}-free buffer was constant over time (Figure 2.13A). Intensities were corrected for autofluorescence, which was also constant and lower than that of dye-loaded cells by a factor of 13 for excitation at 340 nm and a factor of 20 for excitation at 380 nm (Figure 2.13B, inset). Serial dilution and plating of cells confirmed that ~90% of the dye-loaded cells were viable.

2.4.2 Vector construction.

The sequence encoding the adenine riboswitch aptamer domain was inserted between the TΨC and D stem-loops of human $tRNA^{lys3}$ as a synthetic gene in a pIDTSMART vector (IDT). The gene was inserted into pET21a(+) using *Hind* III (5') and *Xba* I (3') cloning sites for expression under the control of a T7 promoter (Figure 2.15). The transcript is a chimeric tRNA in which the anti-codon stem is replaced by the aptamer domain.[30,31]

2.4.3 RNA expression.

The aptamer-tRNA construct (Figure 2.1B) was expressed in BL21(DE3) *E. coli* cells in LB medium at 37 °C. When the OD_{600} reached 0.6, RNA expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (1 mM). Expression was allowed to proceed for 30 min, at which time aliquots were removed and added to 2-aminopurine, 2,6-diaminopurine, or 3-methyladenine dissolved in LB (final concentration 1 mM) or to LB alone. Expression was continued for 1 h, at which time

cell aliquots were subjected to either in-cell or *in vitro* SHAPE.

To measure the concentration of aptamer construct in the cellular RNA, aliquots of total RNA and serial dilutions of purified construct were analyzed by denaturing urea polyacrylamide gel electrophoresis. Gels were stained with SYBR Gold (Invitrogen) and a standard curve was constructed using the band intensities of samples corresponding to the purified construct. The standard curve was used to determine the concentration of the aptamer construct in the total RNA samples. The aptamer construct (~7 g/L) was determined to represent ~7% of the total RNA concentration (100 g/L determined using NanoDrop 1000 Spectrophotometer). To estimate the concentration of construct in cells, we took 7% to the total RNA concentration in *E. coli* (75-120 g/L[49]) to correspond to the aptamer construct. From the molecular weight of the construct (43.4 kD), the concentration of construct in cells is 120-190 μM.

2.4.4 In-cell SHAPE.

The aptamer construct was expressed in the presence or absence of ligand as described above. Following RNA expression, cellular RNA was modified by adding aliquots of cells (1 mL) to 20 μL of 300 mM 1M7 in DMSO or to neat DMSO (2% vol/vol final co-solvent, conditions consistent with normal cell growth, Figure 2.16). The samples were incubated with shaking for 3 min at 37 °C. Total cellular RNA was recovered from protoplasts and isolated as described.[32] RNA was precipitated with ethanol, washed three times with 70% (vol/vol) aqueous ethanol, and resuspended in 15 μL of deionized H_2O.

Primer extension of the aptamer construct was achieved using sequence-

specific primers, containing locked nucleic acid (LNA) nucleotides, and 5' fluorescent labels. Specifically, a 12-nt DNA primer (Exiqon) complementary to the 5' end of the acceptor stem and containing three locked nucleic acid (LNA) nucleotides (5'-*T*G*G*C*G*CCCGAAC-3', where the italic font indicates LNA nucleotides) was used. The primer was labeled with either a 5' 6-FAM or 5-HEX fluorescent dye. The 6-FAM labeled primer was used for the (+) or (–) 1M7 channels; the 5-HEX was used for the sequencing channel. RNA (1.5 µg) and primer (2 pmol) were mixed and diluted to 13 µL with H_2O. The sample was then incubated for 5 min at 65 °C and then on ice for 3 min. The reaction mixture contained 4 µL of 5× SuperScript first-strand buffer, 1 µL 0.1 M dithiothreitol, 1 µL dNTPs (10 mM each), and 1 U SuperScript III (Invitrogen). cDNA products were resolved by capillary electrophoresis.

For sequence analysis, the reactions were prepared as described above, except the RNA was not treated with 1M7 or DMSO, the RNA-primer mixture was diluted to 12 µL, and 1 µL of 10 mM ddATP was added prior to adding SuperScript III. Reactions were incubated at 55 °C for 60 min, and then at 70 °C for 15 min. cDNA was recovered by ethanol precipitation and washed three times with 70% (v/v) aqueous ethanol. Pellets were dried under vacuum and resuspended in 10 µL of deionized formamide. The products were resolved by capillary electrophoresis on an Applied Biosystems 3500 DNA capillary electrophoresis sequencer.

2.4.5 *In vitro* SHAPE.

Following expression, RNA was recovered from protoplasts and isolated as previously described[32] and incubated in 50 mM HEPES-KOH (pH 8.0), 200 mM potassium acetate (pH 8.0) and 1 to 30 mM $MgCl_2$ for 30 min. The aptamer construct

was purified by anion exchange, fast performance liquid chromatography. Ligand (1 mM) was added after folding and incubation was continued for 10 min. The aptamer construct (8 pmol) was then added to 1/50 volume 300 mM 1M7 in DMSO or to neat DMSO. The samples were incubated at 37 °C for 3 min. RNA was precipitated with ethanol, washed three times with 70% (vol/vol) aqueous ethanol, and resuspended in 15 μL deionized H_2O.[32] Primer extension and capillary electrophoresis were used to determine sites of adduct formation, as described for in-cell SHAPE.

2.4.6 SHAPE data analysis.

Electropherograms were analyzed with QuShape software (Figure 2.4).[50] After baseline and mobility shift corrections, peaks in the (+) and (–) 1M7 channels were aligned, and Gaussian integration was used to quantify peaks areas. The SHAPE reactivity is reported as the area of the (+) 1M7 peaks minus the area of the no-reagent background peaks. For correlation analyses (Figure 2.11), we monitored the following single-stranded and terminal base paired nucleotides: A21-U25, U31-U39, A45-C54, G59-C67, and G72-U75.

2.4.7 Calculations for ΔR values shown in Figure 2.12.

Values for long-range tertiary contacts and for nucleotides within the ligand binding pocket represent the average difference in reactivity.

First, the absolute value of the differences (∆R) were calculated using buffer with 1 mM Mg^{2+} as the initial condition:

$$\Delta R_0 = | R_{\text{buffer+1mM Mg}^{2+}} - R_{\text{buffer+1mM Mg}^{2+}}|$$

$$\Delta R_{\text{complete change in buffer}} = |R_{\text{buffer+2AP+1 mM Mg}^{2+}} - R_{\text{buffer+1 mM Mg}^{2+}}|$$

$$\Delta R_{\text{cells}} = |R_{\text{cells}} - R_{\text{buffer+1mM Mg}^{2+}}|$$

$\Delta R_{cells+2AP} = |R_{cells+2AP} - R_{buffer+1\ mM\ Mg^{2+}}|$

$\Delta R_{buffer+10\ mM\ Mg^{2+}} = |R_{buffer+10\ mM\ Mg^{2+}} - R_{buffer+1mM\ Mg^{2+}}|$

ΔR values were not calculated for binding pocket nucleotides A24, G46, and C53 or for long-range contacts C60 and C61 because these nucleotides have low SHAPE reactivities under all conditions, such that the ΔR values were less than the standard deviation from three trials of the free aptamer in buffer plus 1 mM Mg^{2+}.

ΔR values for each nucleotide were then normalized to a scale where 0 is defined as the value for the free aptamer in buffer at 1 mM Mg^{2+} and 1.0 represents the 2AP-bound state in buffer at 1 mM Mg^{2+}. The normalized values were then grouped by region and averaged. The average value for each region is a measure of organization relative to the fully folded, ligand-bound state in buffer at 1 mM Mg^{2+}.

Region J1-2 in the free aptamer in cells is unique because values of $R_{cells} - R_{buffer+1mM\ Mg^{2+}}$ are opposite in sign to $R_{buffer+2AP+1\ mM\ Mg^{2+}} - R_{buffer+1\ mM\ Mg^{2+}}$. This situation arises because the reactivity of nucleotides in J1-2 increase in cells relative to reactivities in buffer with 1 mM Mg^{2+}(Figure 2.5), but decreases with ligand binding (Figure 2.7A). Therefore, we multiplied these values by -1 to reflect that ΔR is opposite in sign to that of the ligand-bound state in buffer.

2.5 Figures

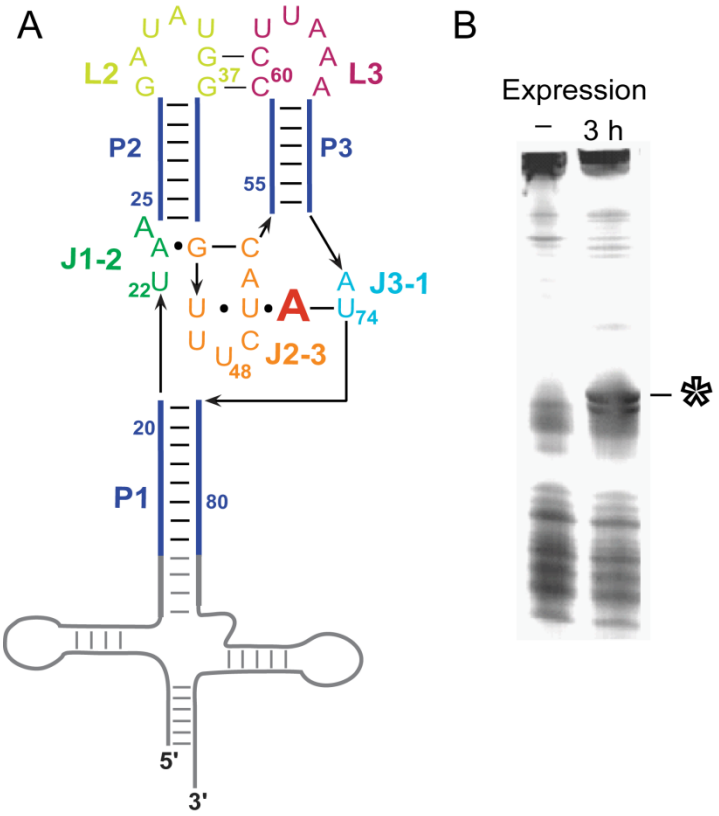

Figure 2.1 Adenine aptamer domain expressed in *E. coli* using a tRNA scaffold. (A) Structure of the aptamer-tRNA construct. The adenine ligand is shown as a red "A". Numbering corresponds to the *add* aptamer domain.[13] (B) Denaturing polyacrylamide gel electrophoresis of total cellular RNA 3 h after induction of aptamer expression. The 143-nt aptamer-tRNA chimera product is marked with an asterisk.

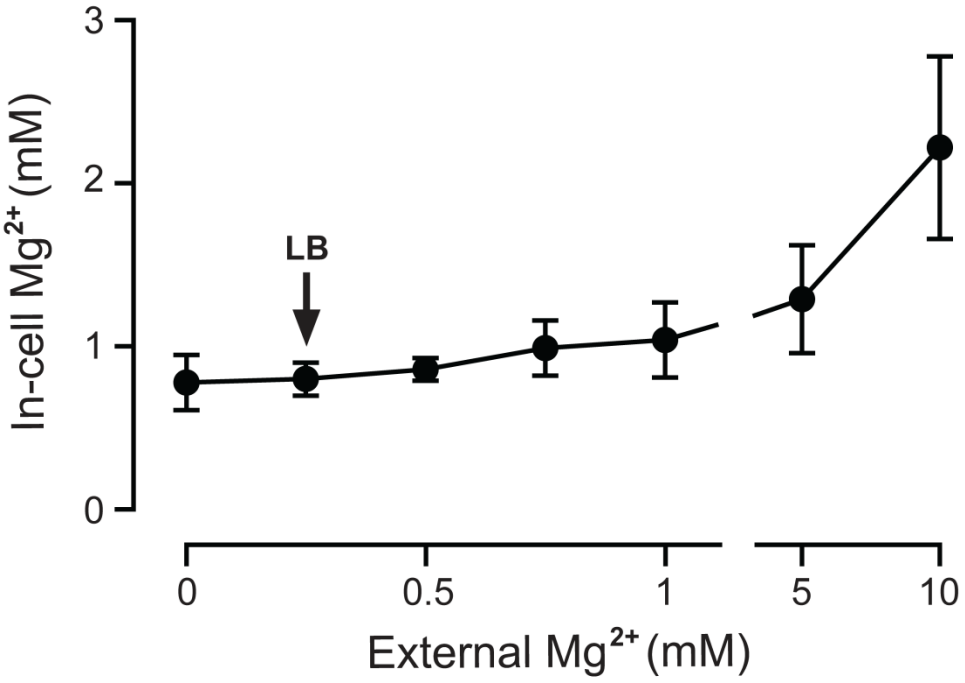

Figure 2.2 Free Mg^{2+} concentration in *E. coli* cells. Free Mg^{2+} was measured using the ion-selective fluorophore, mag-fura-2. Free Mg^{2+} in *E. coli*, measured at late-log phase, is 0.8 ± 0.2 mM in standard LB media at 37 °C; the external concentration was 0.25 mM (arrow). Error bars show the standard deviation of the mean from three trials.

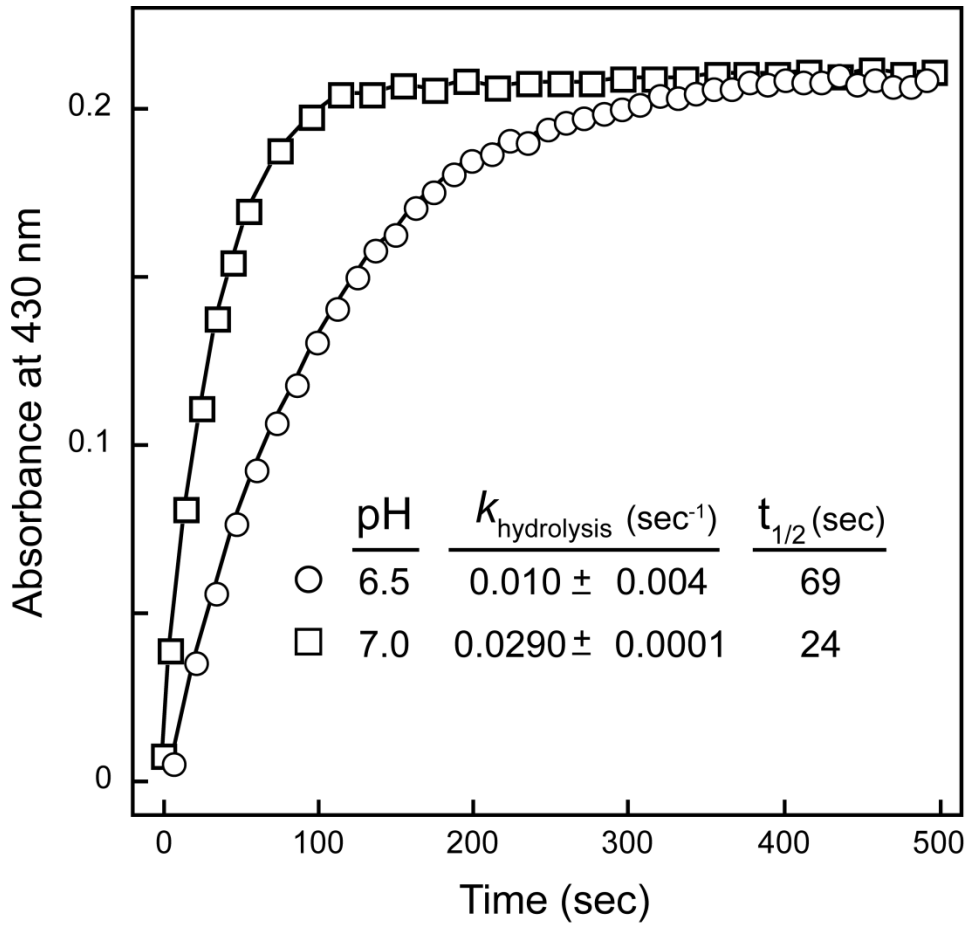

Figure 2.3 Pseudo-first order hydrolysis of 1M7 in LB at 37 °C at pH values of 6.5 and 7.0, which correspond to that of the culture media at the beginning and in the late-log phases of cell growth. Reagent hydrolysis was measured in a cuvette at 37 °C by adding 100 μL of 2.0 mM 1M7 in DMSO to 900 μL of equilibrated LB (pH 6.5 or 7.0). Pseudo-first-order rate constants were obtained by monitoring the 430 nm absorbance of the hydrolysis product (2-methylamino-4-nitrobenzoate).

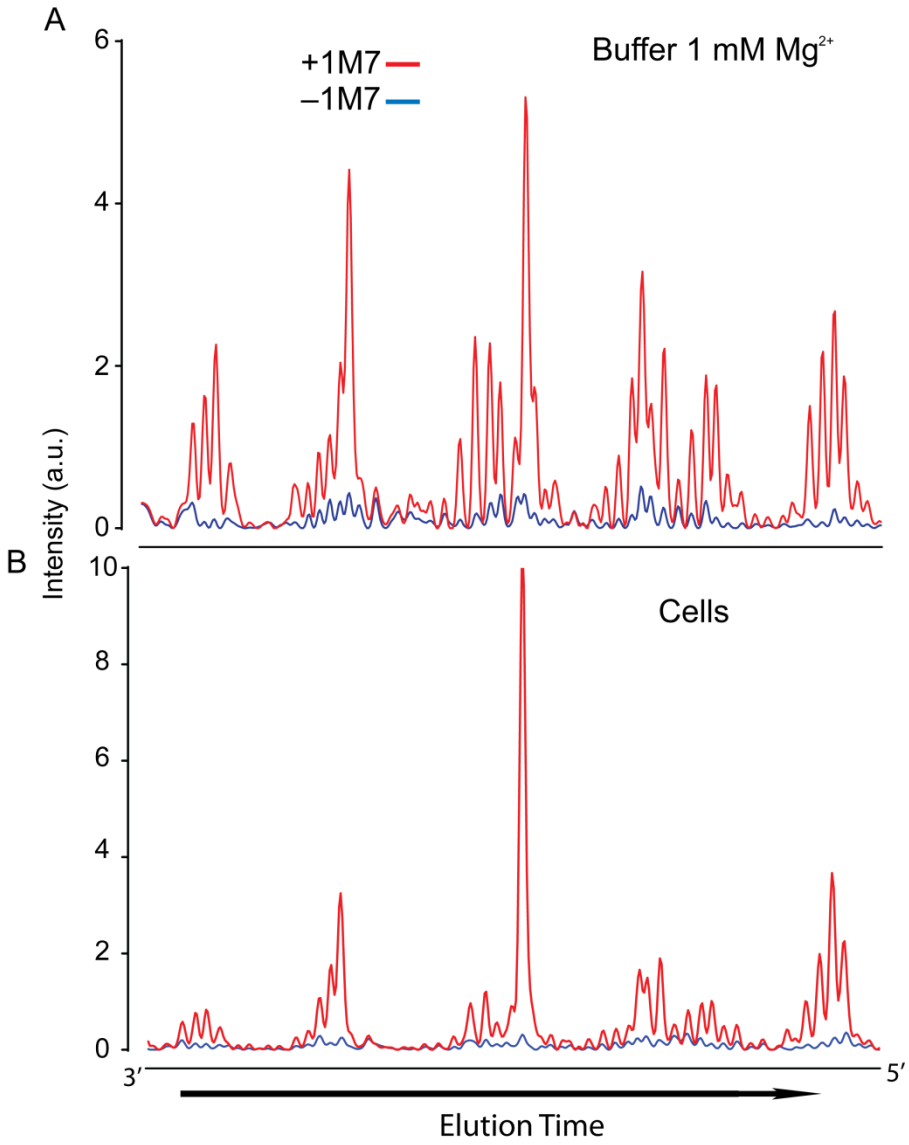

Figure 2.4 Electropherograms from SHAPE analysis of the adenine aptamer (A) in buffer with 1 mM Mg^{2+} and (B) in cells. Signal decay and baseline corrections were applied using qSHAPE software.[50]

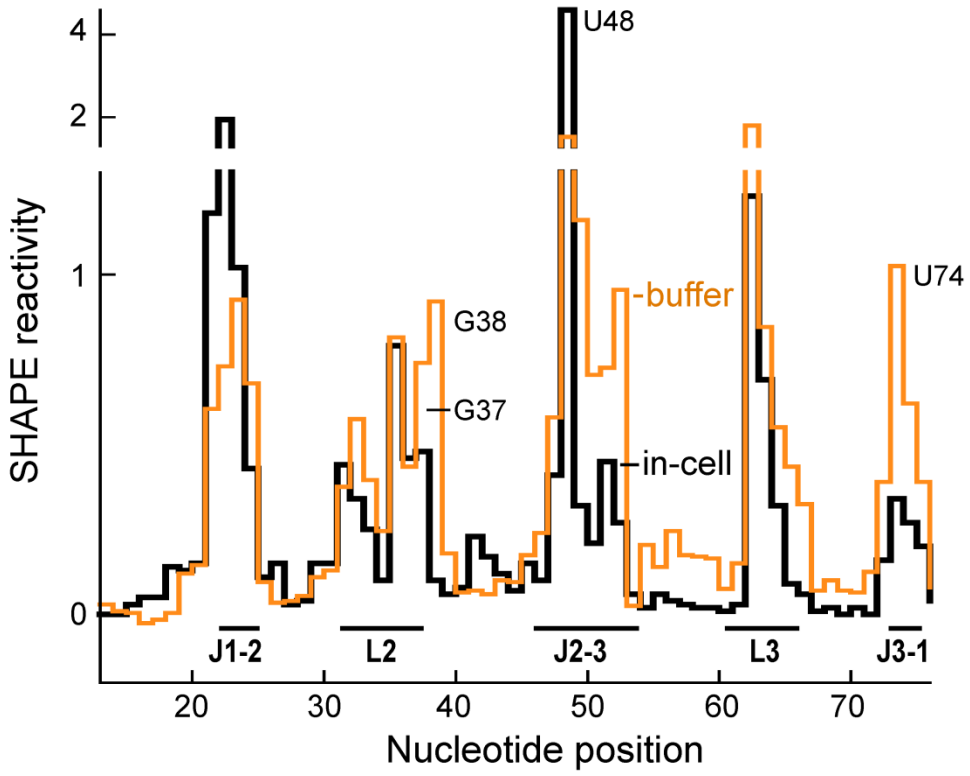

Figure 2.5. SHAPE reactivity profiles of the adenine aptamer domain in cells (black) and in buffer at a near-physiological Mg^{2+} (1 mM) concentration (orange).

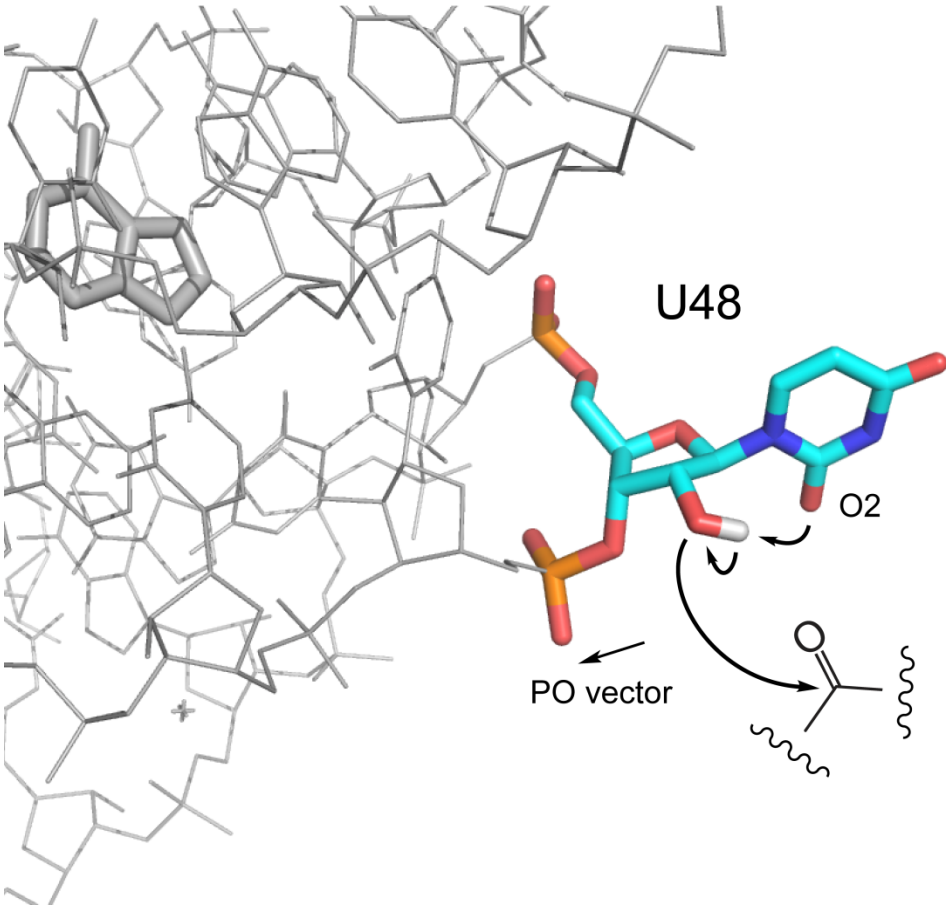

Figure 2.6 Mechanism of SHAPE hyper-reactivity at nucleotide U48. The conformation encompasses two features that facilitate SHAPE reactivity. First, the anionic 3'-phosphate diester is directed away from the reactive 2'-hydroxyl group (illustrated as the PO vector). Second, the 2-keto group of the uracil moiety is positioned to function as a general base and lower the pK_a of the attacking 2'-OH group.[44]

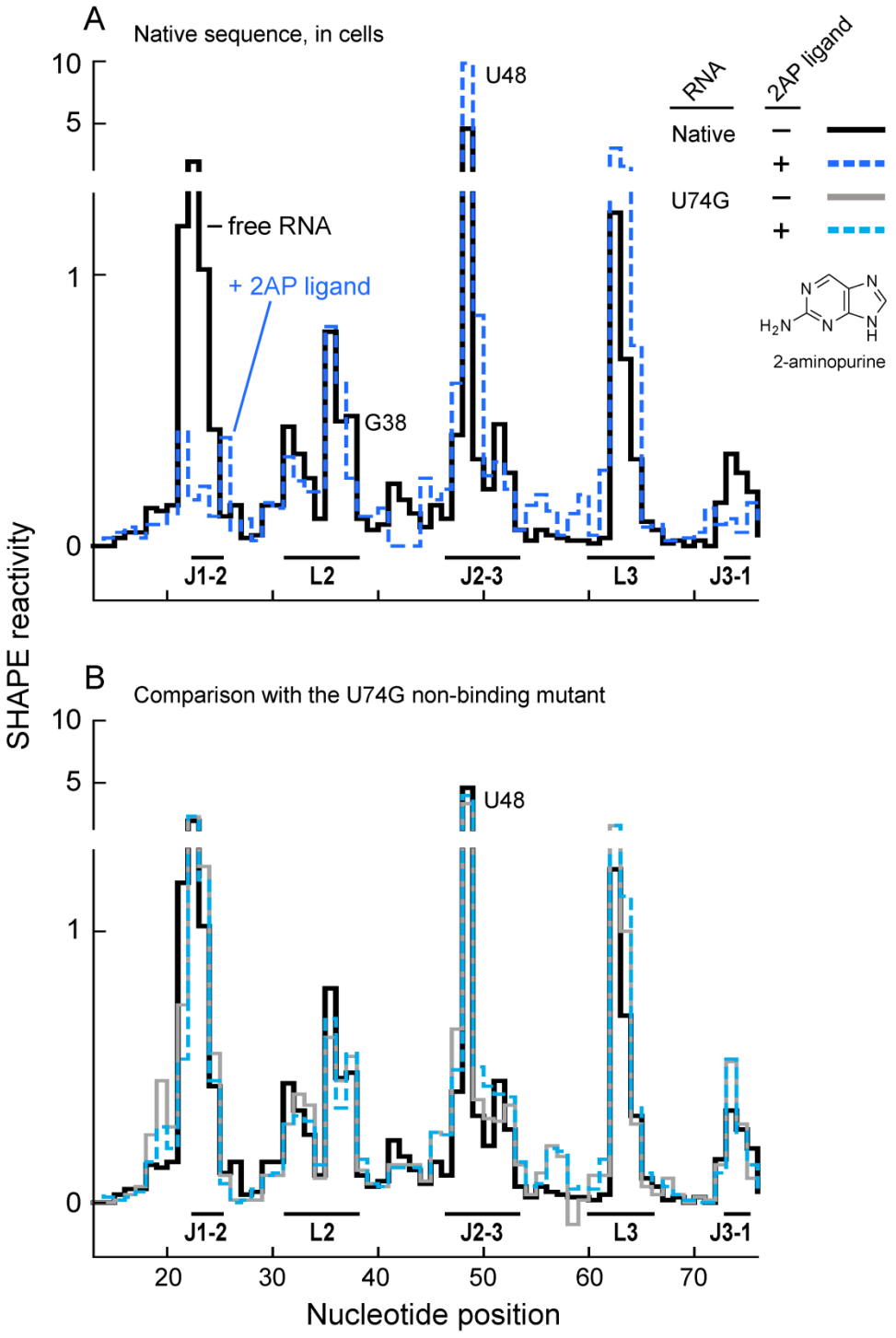

Figure 2.7 In-cell SHAPE reactivity and ligand binding by (A) the native sequence adenine aptamer and (B) the non-binding U74G mutant. When present, the 2AP ligand concentration was 1 mM.

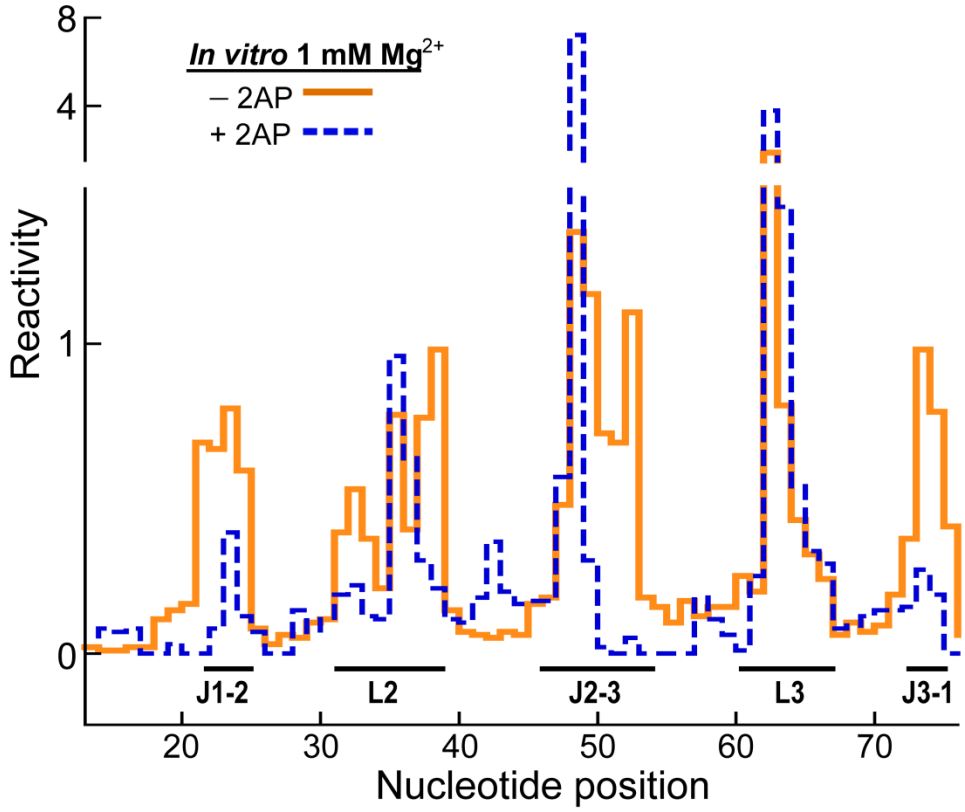

Figure 2.8 *In vitro*, the aptamer undergoes a substantial conformational change upon adding the 2AP ligand. SHAPE reactivity profiles for the ligand-free and ligand-bound aptamer are shown.

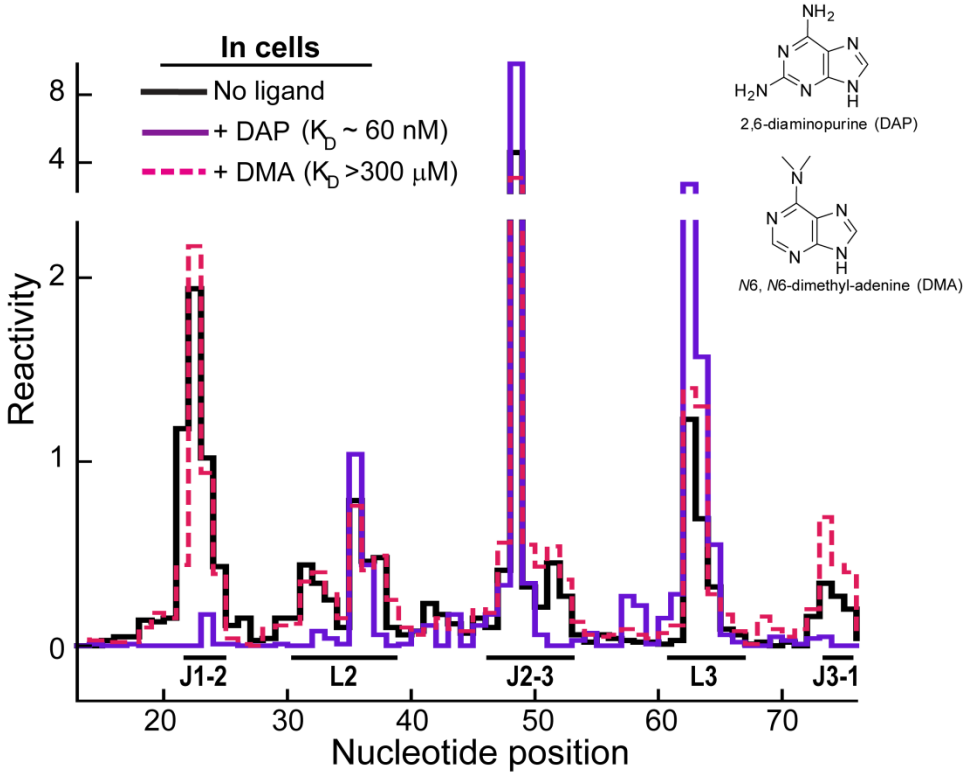

Figure 2.9 In cells the aptamer binds 2,6-diaminopurine (DAP) but not *N*6,*N*6-dimethyl-adenine (DMA) ligands. SHAPE reactivity in cells without ligand and with 1 mM DAP or DMA ligand.

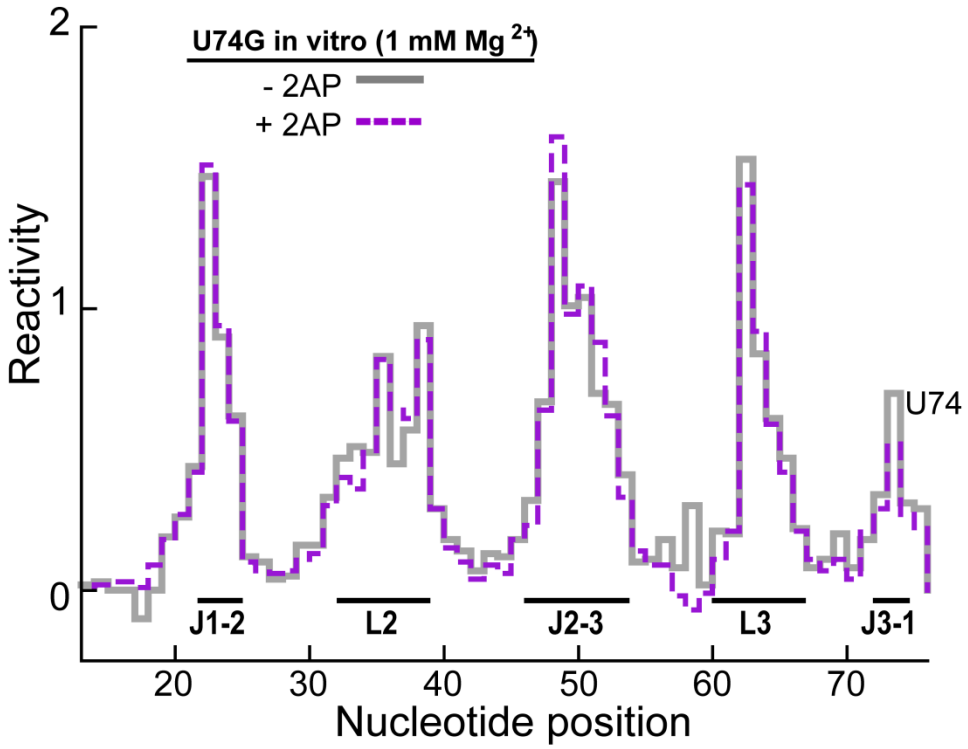

Figure 2.10 U74G mutant does not bind ligand *in vitro*. SHAPE reactivities for the native sequence aptamer and the U74G mutant in the presence and absence of the 2AP ligand.

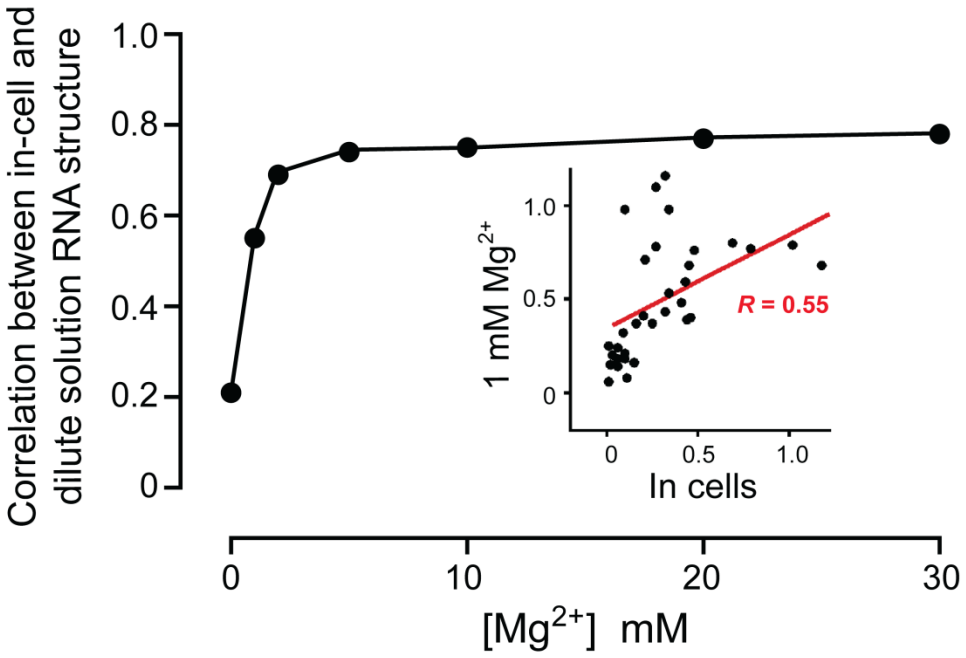

Figure 2.11 Correlation between in-cell and *in vitro* SHAPE reactivities for the ligand-free aptamer domain RNA as a function of Mg^{2+} concentration. Data are from single-stranded and terminal base-paired nucleotides. *Inset*: *In vitro* SHAPE reactivities at 1 mM Mg^{2+} plotted against in-cell reactivities.

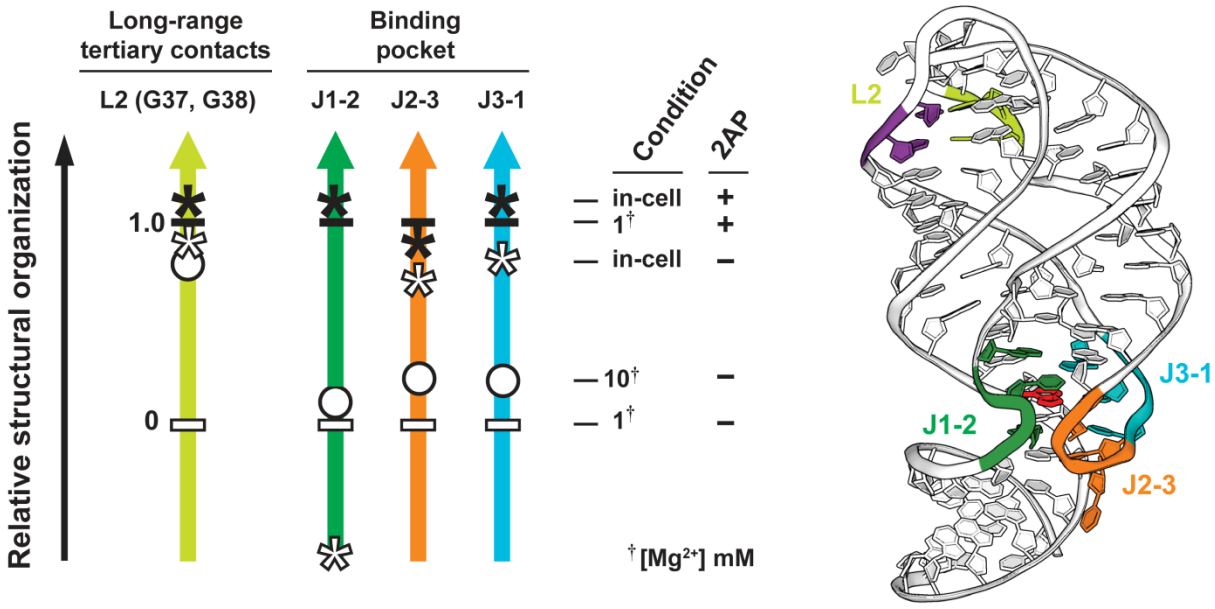

Figure 2.12 The adenine aptamer is pre-organized in cells. Relative levels of organization were calculated for nucleotides in each of four structural elements using a scale in which the reactivity of the nucleotide in the free aptamer is defined as 0.0 and that of the nucleotide in the 2AP ligand-bound aptamer in buffer at 1 mM Mg^{2+} is 1.0. Fractional changes were grouped by region and averaged. Arrows are colored using the scheme shown in Figure 2.1. The diameter of the symbols represents the average standard error of the mean for each region (± ~0.1).

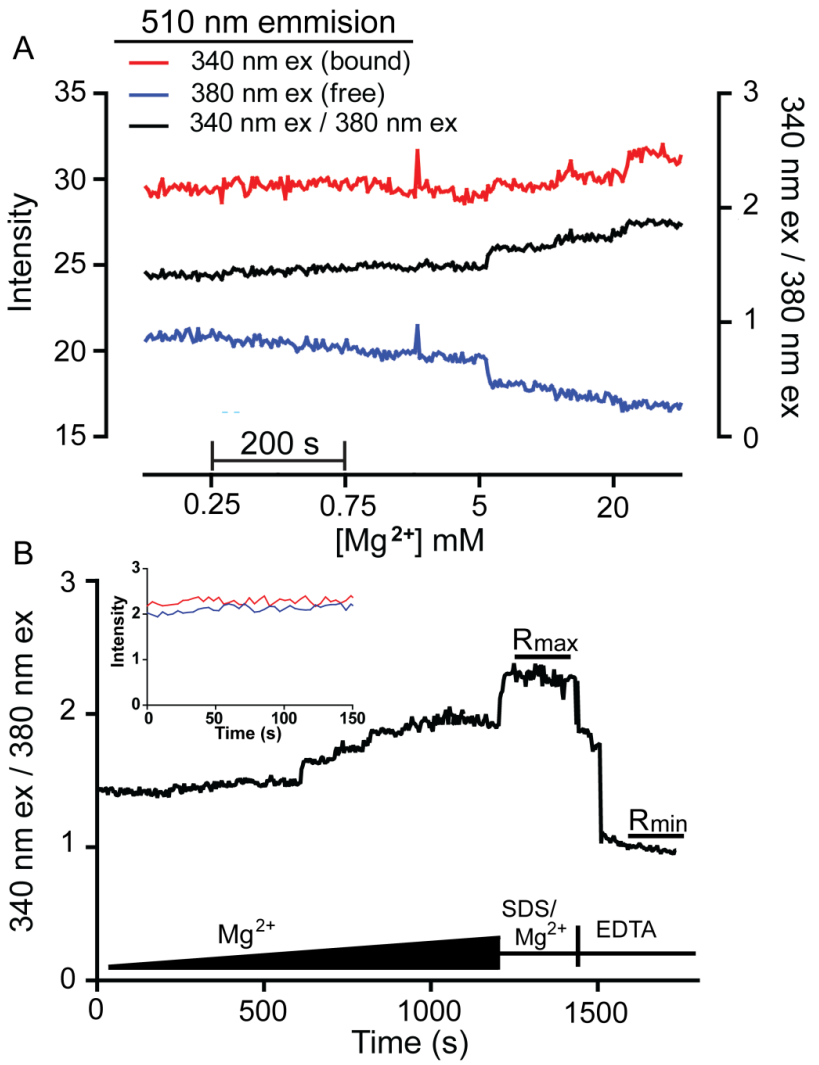

Figure 2.13 Free intracellular Mg^{2+} concentration in *E. coli* cells measured ratiometrically using mag-fura 2 is <1 mM. (A) Fluorescence intensities at 510 nm with excitation at 340 nm (red) and 380 nm (blue), corresponding to the bound and free dye, respectively, were measured as a function of added Mg^{2+}. The ratios (black) are used in Eqn. 1 to obtain the concentration of free Mg^{2+} in cells.[36] (B) Ratio of 510 nm emission intensities from 340 nm and 380 nm excitation for a complete experiment. Initial portion of the ratio plot (black) is the Mg^{2+} titration shown in (A). R_{max} was measured by adding SDS (0.1% m/v final concentration) and additional Mg^{2+} (30 mM final concentration) to the cells. R_{min} was measured by adding excess EDTA to chelate the Mg^{2+}. R_{min} and R_{max} were consistent across all experiments (R_{max} = 2.02 ± 0.20, R_{min} = 0.95 ± 0.05).

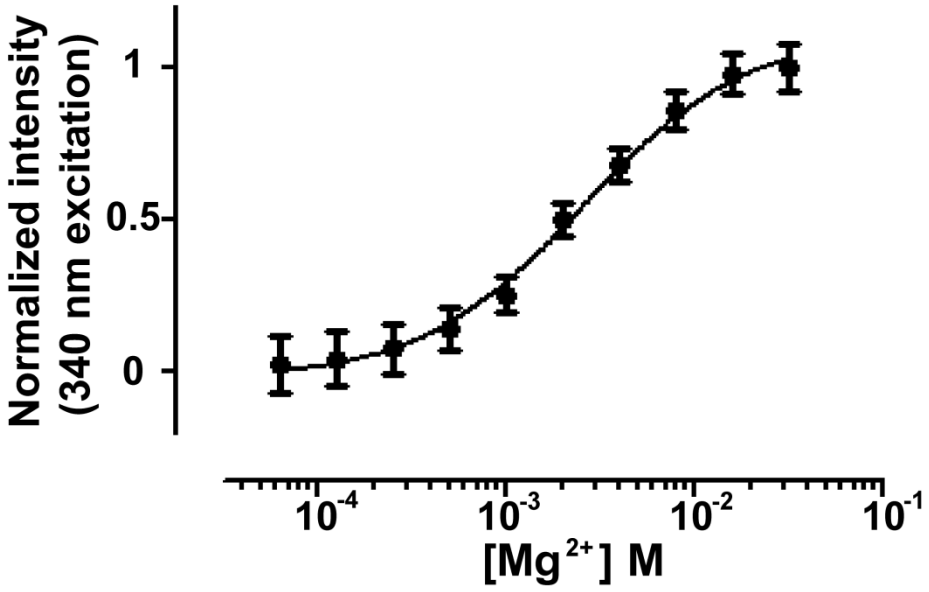

Figure 2.14 Determination of the K_D of the mag-fura 2- Mg^{2+} complex. Normalized intensities at 510 nm (340 nm excitation) were measured in buffer (see Methods) as a function of Mg^{2+} concentration. Intensities were fit to a single-site binding equation to yield the transition midpoint, corresponding to a K_D of 2.5 mM.

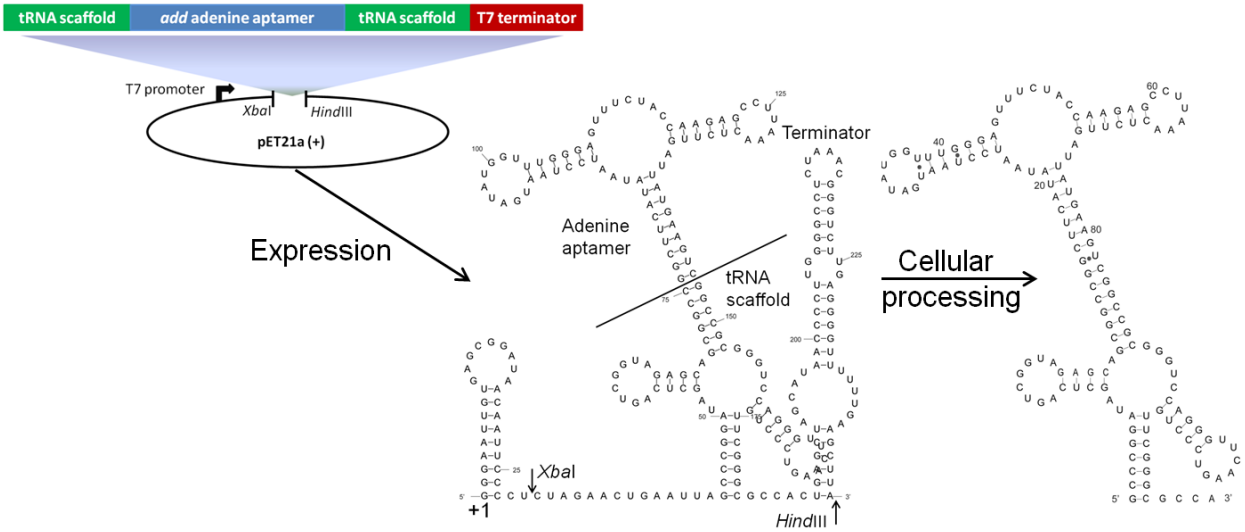

Figure 2.15 Chimeric tRNA-aptamer domain vector design and expression.

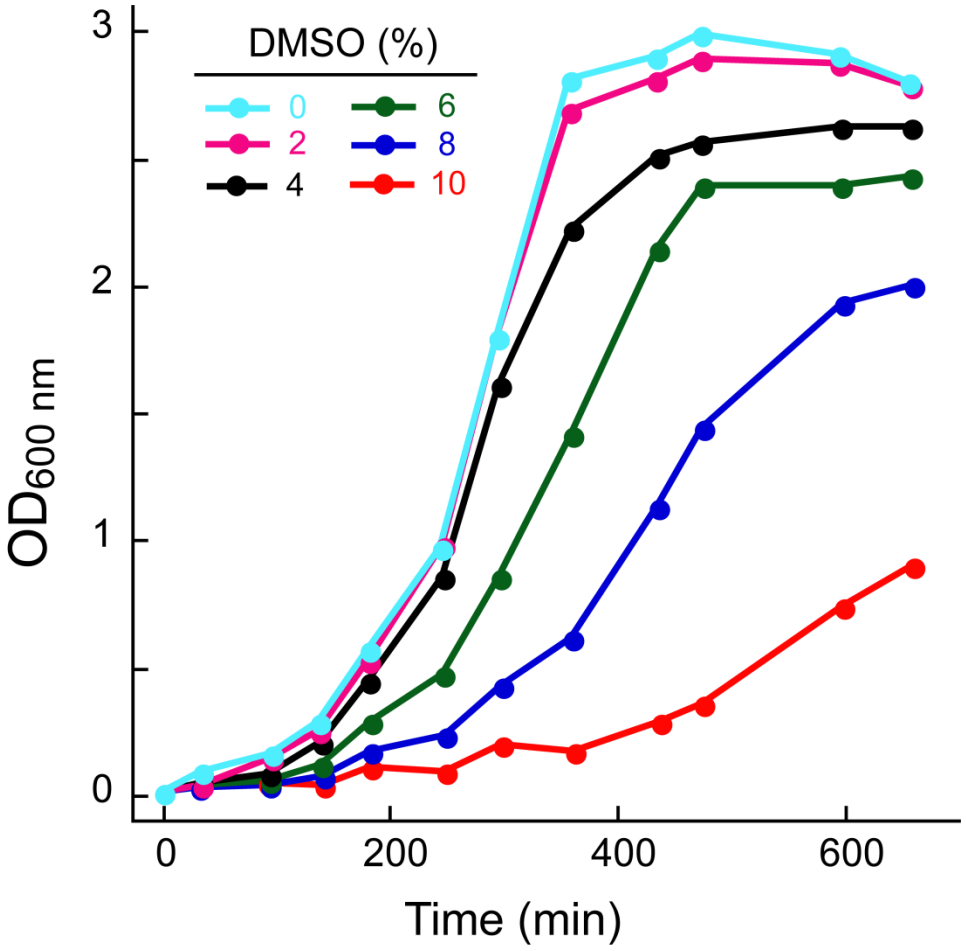

Figure 2.16 Effect of DMSO on *E. coli* growth rate. DMSO concentrations below 4% (v/v) do not significantly perturb cell viability. A final DMSO concentration of 2% (v/v) was using for in-cell SHAPE in this work. DH5α cells (500 mL) were grown in LB to an OD_{600} of 0.3 starting from 5 mL of an overnight culture. Cell aliquots (10 mL) were then added to 90 mL of LB containing 0–10% (v/v) DMSO. Samples were withdrawn at intervals, and the OD_{600} was measured to assess cell growth. The resulting growth curves showed that *E. coli* cells are fully viable at DMSO concentrations below 4% (vol/vol) but growth is compromised at 10% co-solvent.

Chapter 3: Effects of crowding on structure and ligand binding affinity *in vitro*

3.1 Introduction

RNA molecules function in cells where the concentration of macromolecules reaches up to 400 g/L, resulting in at least 40% of the cell volume being occupied.[1] Most studies of RNA structure and function, however, are carried out in dilute buffer. It has been shown that macromolecular crowding *in vitro* can have significant effects on RNA structure.[6,7,9,10] For many biological RNAs, structure is intimately linked to function, which suggests that crowding may affect many biological functions. To fully understand structure and function relationships, it may be important to consider the effects that the crowded intracellular environment may have on RNA.

In this dissertation, I have reported a nucleotide-level analysis of the *add* A riboswitch aptamer domain in healthy, growing *E. coli* cells (Chapter 2). The *add* A riboswitch aptamer is highly conserved and its function depends on a conformational change upon specific recognition of adenine in cells, making it a good candidate for RNA structure function studies (Figure 1.1).[12,13] Ultimately, I have shown that the conformation of the free aptamer in cells is different than that observed in dilute buffer with physiologically relevant Mg^{2+} concentration (Figure 2.2), and these effects are not due to endogenous ligand binding (Figures 2.7 and 2.11). On the other hand, the cellular environment has almost no effect on the ligand-bound

aptamer, which is highly organized and compact both in cells and in dilute buffer, in agreement with NMR analysis[16] and the x-ray crystal structure[13] of the ligand-bound aptamer.

Crowding effects can arise from two phenomena: hard-core repulsions and non-specific chemical interactions.[2,3] Hard-core repulsions arise from excluded volume and reduce conformational space, therefore favoring compaction. Non-specific chemical interactions include repulsive or attractive charge interactions.[27,28] Repulsive interactions arise from interactions between like charges and reinforce the hard-core effect, favoring compact states. Attractive interactions include hydrogen-bonding, juxtaposed complementary charges, and hydrophobic contacts. Attractive interactions that expose more interacting surface to the crowding molecules destabilize folded, compact states.[27]

The observed conformation of the aptamer in cells was more compact than in dilute buffer, indicated by the stabilization of L2 and L3 interactions.[26] This observation suggests that, overall, the hard-core effect and/or repulsive charge interactions may dominate in cells. However, region J1-2 was destabilized in cells, suggesting that attractive chemical interactions may also play a role in effecting the in-cell aptamer structure. Elucidating the specific effects of crowding in the cellular environment may help to establish *in vitro* conditions to study RNA that are more physiologically relevant than dilute buffer with high Mg^{2+}.

In this chapter, I investigate the effects of non-specific RNA-RNA interactions on the free aptamer structure *in vitro* using SHAPE. It was expected that crowding the aptamer with RNA will introduce both repulsive interactions, due to the negatively

charged backbone, and attractive interactions, arising from the potential for inter-molecular base-paring. Observed structural effects on the free aptamer crowded by high concentrations of cellular RNA, suggest that repulsions dominate. The SHAPE profile for the aptamer in this mixture, however, indicates that the binding pocket is not affected by RNA-RNA interactions in the same way as it is by the cellular environment.

The effects of the commonly used synthetic crowding agent, polyethylene glycol (PEG), on both the ligand-free and ligand bound aptamer, were also investigated. Previous results show that favorable interactions between PEG and DNA increase with decreasing PEG size,[51] suggesting that crowding with PEG may present an opportunity to separate effects of non-specific attractive interactions and hard-core repulsion on aptamer structure. Similar results were obtained with low molecular weight PEGs (> 3350 Da) as observed with RNA-RNA interactions, in that SHAPE profiles correlated with the in-cell profile better that obtained in dilute buffer, but the reactivity of the ligand-binding pocket was unlike that observed in cells. For small molecular weight PEGs, the results suggest that the aptamer structure is strongly affected by favorable interactions, in agreement with observations involving DNA.[51] For ethylene glycol (EG), favorable interactions also affected the ligand bound aptamer. However, no significant differences were observed for the ligand-bound aptamer in the presence of PEG > 3350 Da. Binding studies for the aptamer in the presence PEG were also performed. The results showed that the structural effects observed for the aptamer in PEG had little effect on binding affinity, however a small, size-dependent trend was observed. These results suggest that, while hard-

core repulsions induce a structure more like the in-cell structure for the free aptamer, the effects of the cellular environment are more complex than can be mimicked by a simple crowded solution. Additionally, binding studies suggest that the binding affinity is not strongly affected by structural changes induced by PEG.

3.2 Results

3.2.1 Effect of RNA-RNA interactions on ligand-free aptamer structure

To examine whether non-specific RNA-RNA interactions can explain the differences between the purified aptamer *in vitro* (1 mM Mg^{2+}) and the in-cell results (Chapter 2) I probed the structure of the aptamer construct in the presence of total cellular RNA (Figure 3.1). The SHAPE profile for the ligand-free aptamer in the presence of total *E. coli* RNA (400 ng/μL) had high or moderate reactivity in the single-stranded binding pocket regions, and low reactivity in helices, consistent with previous observations in which the ligand-free aptamer has a disordered binding pocket and fully intact helices.[12,16,26,34] SHAPE profiles for the pure aptamer in the absence and presence of total RNA correlated with an R-value of 0.7 (Figure 3.1A). Effects of total cellular RNA on aptamer SHAPE profiles were not significantly different when the concentration was reduced to 100 ng/μL (R=0.7, data not shown). For the aptamer with added cellular RNA, nucleotides in L3 and J2-3 exhibited small decreases in reactivity, and G38 in L2, which is highly reactive for the pure aptamer, was unreactive. These observations would suggest stabilization of tertiary structure, particularly due to interactions between L2 and L3. However, G37, which pairs with C61 in L3 when L2-L3 interactions are stabilized, is increased by ~0.8 SHAPE units.

This is inconsistent with significant stabilization of the L2-L3 interaction. Taken together, these observations suggest that while there is modest tertiary stabilization, likely reflecting inter-RNA repulsion involving negatively-charged RNA backbones, favorable intermolecular interactions may play a role.

The aptamer structure effected by RNA-RNA interactions *in vitro* is more similar to that observed in cells for the aptamer (data from Chapter 2) compared to the pure aptamer (R=0.8, Figure 3.1B). The increased correlation can be attributed to better agreement with the in-cell profile in regions J2-3 and L3 compared to pure aptamer. However, for the aptamer with added RNA, J3-1 is significantly higher in reactivity and J2-1 is lower in reactivity compared to cells. Additionally, the aptamer structure induced by RNA-RNA interactions has low reactivity for nucleotide G38 but exhibits high reactivity for nucleotide G37, while the in-cell structure has low reactivity for both, suggesting that L2-L3 interactions are significantly more stable in cells. In conclusion, the conformation of the aptamer induced by RNA-RNA interactions agrees better with that observed in cells compared to the purified aptamer. However, this treatment fails to fully recapitulate the in-cell state.

3.2.2 Structural effects of PEG

3.2.2.1 Effect of PEG on the ligand-free aptamer

The effects of PEG on aptamer structure were examined using SHAPE. The aptamer was probed in the presence of ethylene glycol (EG), PEG 3350, 12k, and 35k at 100 g/L. Compared to the ligand-free aptamer in dilute buffer, nucleotide reactivities in the binding pocket are significantly reduced in the presence of EG (Figure 3.2). This observation suggests that EG strongly interacts with the aptamer,

and either decreases nucleotide dynamics or causes misfolding. EG also causes increased reactivity in the P1 helix, suggesting P1 is destabilized. PEG 3350 induces a structure that exhibits both hard-core effects, as indicated by low reactivity in L2 and L3, and chemical interactions, indicated by slight disruption of the P1 stem (Figure 3.2). Binding pocket and P1 helix reactivities for the free aptamer converge with those in dilute buffer when PEG size is increased to 12k. The reactivities in L2 and L3 are significantly reduced with the addition of PEG 35k, compared to dilute buffer. This observation suggests that the tertiary interactions between L2 and L3 are significantly more stable in the presence of PEG 35k, which is consistent with compaction induced by hard-core repulsions. In sum, increasing PEG size reduces the chemical interaction effect but increases the effect of hard-core repulsions on the ligand-free aptamer.

To compare the effects of PEG on the ligand-free aptamer structure to the effects induced by the cellular environment, Pearson-Correlation coefficients (R) were calculated between in-cell and *in vitro* (with PEG) SHAPE reactivities and plotted as a function of PEG size (Figure 3.3A). Compared to the correlation between in cells and dilute buffer with 1 mM Mg^{2+}(R=0.6), adding EG to the aptamer reduced the correlation to 0.1. The correlation value increased with increasing PEG size, surpassing the dilute buffer correlation value when PEG > 3350. The correlation increased to 0.8 with the addition of PEG 35k.

Directly comparing the free aptamer structure in cells and in the presence of PEG 35k, shows that, similar to results with increasing Mg^{2+} concentrations (Figure 2.5), and with RNA-RNA interactions (Figure 3.1), the improved correlation,

compared to that in dilute buffer with 1 mM Mg^{2+}, is largely due to stabilization of L2-L3 interactions (Figure 3.2). However, the effects of the cellular environment on the aptamer binding pocket are unique. These results confirm that the free aptamer conformation in cells is dominated by hard-core repulsions and repulsive charge interactions but attractive chemical interactions must play a role. The combination of hard-core repulsions and chemical interactions that induce the in-cell aptamer structure are not mimicked by PEG in this study.

3.2.2.2 Effect of PEG on the ligand-bound aptamer

The reactivity patterns were nearly identical for the ligand-bound aptamer in cells, in dilute buffer, and in PEG 3350 or higher were nearly identical (data not shown). The correlation coefficient reached 0.9 for PEG 35k (Figure 3.3B). However, EG gave a low correlation with the in-cell ligand bound structure (R~ 0.4). EG significantly disrupts the ligand-bound structure compared to in cells (nearly identical to dilute buffer R ~1.0) (Figure 3.4). Specifically, reactivities in helices P1 and P2 are much higher than those observed for the ligand-bound structure in cells, suggesting that these helices are disrupted by EG. In summary, although the ligand-bound structure is not significantly affected by high molecular weight PEGs, the interactions between the aptamer and EG are strong enough to disrupt the highly structured, ligand-bound state. Considering that the ligand-bound structure in cells is identical to that in dilute buffer, these results suggest that strong chemical interactions with the aptamer, as observed with EG, are not present in cells or are overcome by hard-core repulsions.

3.2.3 Effect of PEG on 2AP affinity

3.2.3.1 Affinity for 2AP in buffer with 1 mM Mg^{2+} ($K_{Dapp,dilute}$)

The equilibrium dissociation constant for the aptamer-2AP complex in dilute buffer with 1 mM Mg^{2+} was determined by monitoring the fluorescence quenching of 2AP upon titration of aptamer (Figure 3.5). From three experiments, $K_{Dapp,dilute}$ was 48 ± 5 nM, This is in fairly good agreement with the reported K_{Dapp} (~100 nm), which was determined at a lower temperature and in different buffer conditions.[45]

3.2.3.2 Size-dependent effects of PEG on aptamer binding affinity

To assess effects of crowding on aptamer binding affinity ($K_{Dapp,\ crowd}$) measurements were carried out in the presence of the EG, PEG 200, 400, 1000, 3350, 8k, and 12k (numerical value is the average molecular weight in Da). While the effects of PEG on K_{Dapp} for the aptamer-2AP complex were not large, there was clear size dependent trend (Figure 3.6). EG had the largest effect on K_{Dapp}, increasing $K_{Dapp,\ dilute}$ three fold. Despite the structural differences observed for the free and bound aptamer in the presence of EG, 2AP fluorescence quenching by RNA was comparable to that in buffer alone , confirming that the aptamer binds 2AP in the presence of EG (data not shown).

As PEG size was increased, $K_{Dapp,crowd}$ decreased and got lower than $K_{Dapp,\ dilute}$ (Figure 3.6). $K_{Dapp,\ crowd}$ was less than $K_{Dapp,\ dilute}$ in the presence of PEG 12k. These results suggest that structural effects brought about by chemical interactions with small PEGs, result in slightly decreased binding affinity. As hard-core repulsions become dominant with increasing PEG size, the binding affinity is slightly increased compared with that observed in dilute buffer.

3.3 Discussion

The cellular environment is crowded. I have shown that this crowding significantly perturbs the structure of the adenine aptamer domain in cells, indicating that it is important to consider these effects when exploring RNA structure and function relationships. The cellular environment, however, is complex and difficult to mimic *in vitro*. Effects of crowding in the cellular environment likely arise from both hard-core repulsions and non-specific chemical interactions. To better understand these effects, the aptamer structure was probed using SHAPE in the presence of total cellular RNA and PEG, both of which impose hard-core repulsions and non-specific chemical interactions. Binding studies were carried out to determine the effects of structural changes induced by PEG on ligand binding affinity for the aptamer.

For the ligand-free aptamer, correlations with the structure observed in cells increased from 0.6 in buffer to 0.8with the addition of total RNA. However, key differences were observed that suggest that tertiary interactions are more stable in cells, and that the binding pocket is stabilized in a unique way by the intracellular environment (Figure 3.1).

Addition of PEG resulted in size dependent effects. EG induced large effects on both the free and bound aptamer, which suggests that there are strong chemical interactions between EG and the aptamer (Figures 3.2 and 3.4). These effects decreased with increasing PEG molecular weight. Compared to dilute buffer, the structure of the free aptamer, in the presence of high molecular weight PEGs is more compact, and L2-L3 interactions are at least partially in place. The L2-L3

interactions became increasingly stable with increased PEG size, suggesting the aptamer was further compacted, and that hard-core repulsions are the dominant phenomenon for high molecular weight PEGs. However, the binding pocket of the ligand-free aptamer remained highly reactive even with high molecular weight PEGs, which is in contrast to the observed in-cell structure (Figure 3.2), suggesting that crowding RNA with synthetic polymers, such as PEG, does not yield physiologically relevant information.

Binding studies showed that the structural effects induced by PEG have little affect on ligand binding affinity, even for EG, which significantly disrupted both the free and bound aptamer. However, as PEG size decreases, binding affinity also decreases. This suggests that the effects of chemical interactions on the aptamer structure have a small negative effect on binding affinity.

In conclusion, hard-core repulsions dominate in determining the aptamer structure in the presence of high concentrations of total RNA and high molecular weight PEGs. Excluded volume induced by these crowders did result in a more compact state that correlated better with the conformation observed in cells compared to dilute buffer. However, the in-cell aptamer structure was not fully recapitulated in the presence of crowders *in vitro.* Our results suggest that the cellular environment has a unique effect on the aptamer, where there is likely interplay between excluded volume and chemical interactions, the complexity of which is difficult to simulate *in vitro*.

3.4 Materials and methods

3.4.1 Large-scale preparation of the *add* A aptamer

The sequence encoding the adenine riboswitch aptamer domain was inserted between the TΨC and D stem-loops of human $tRNA^{lys3}$ as a synthetic gene in a pIDTSMART vector (IDT) (Figure 2.5). The gene was inserted into pET21a(+) using *Hind* III (5') and *Xba* I (3') cloning sites for expression under the control of a T7 promoter (Figure 2.5). The transcript is a chimeric tRNA in which the anti-codon stem is replaced by the aptamer domain.[30,31]

The aptamer-tRNA construct (Figure 2.1B) was expressed in BL21(DE3) *E. coli* cells in LB medium at 37 °C. When the OD_{600} reached 0.6, RNA expression was induced by addingisopropyl β-D-1-thiogalactopyranoside (1 mM). Expression was allowed to proceed for 3 h. Following expression, total cellular RNA was recovered from protoplasts and isolated as described.[32] The aptamer construct was purified by anion exchange, fast performance liquid chromatography. Purity was assessed by using urea polyacrylamide gel electrophoresis (PAGE).

3.4.2 *In vitro* SHAPE with synthetic crowding agents or total cellular RNA

The purified aptamer construct was folded by incubation in 50 mM HEPES (pH 8.0), 200 mM potassium acetate (pH 8.0), and 1 mM $MgCl_2$ for 30 min at 37 °C. Synthetic crowding agents (100 g/L) or total cellular RNA (100 g/L, 200 g/L, or 400 g/L, isolated from cells without aptamer construct expression as described in section 2.4.5) were then added. The final concentration of aptamer construct was 1 μM. Incubation was continued for 20 min. For SHAPE of the ligand-bound aptamer, 2AP (1 mM) was added after folding, and incubation was continued for 10 min. The

aptamer construct (10 pmol) was then added to 1/50 volume 300 mM 1M7 in DMSO or to neat DMSO. The samples were incubated at 37 °C for 3 min. RNA was precipitated with ethanol, washed three times with 70% (vol/vol) aqueous ethanol, and resuspended in 15 µL deionized H_2O.[32] Primer extension and capillary electrophoresis were used to determine sites of adduct formation, as described (Chapter 2).

3.4.3 Measuring K_D using 2AP fluorescence

The quenching of 2AP fluorescence intensity was measured at a fixed concentration of 2AP upon addition of RNA, as described,[45] with some modification. Purified aptamer was folded and incubated with co-solutes, as described above. Samples (50 µL) were prepared with no aptamer or aptamer concentrations ranging from 10 nM to 3 µM, containing 0 g/L or 100 g/L crowding agent, 40 nM 2AP in 50 mM HEPES-KOH (pH 8.0), 200 mM potassium acetate (pH 8.0), and 1 mM $MgCl_2$. Fluorescence intensities were measured in a 40 µL cuvette at 25 °C using a Varian Cary Eclipse fluorimeter. Data were collected from 330 nm to 450 nm emission with an excitation wavelength of 300 nm.

Data were processed and binding curves were obtained as described.[45] The data collected between 330 nm and 450 nm were integrated using Graphpad Prism 6 software and fit to a single-site binding equation:

$$dF/F = (1-a)\,[RNA] / K_{Dapp} + [RNA]) \quad (1)$$

where dF is the change in fluorescence intensity, F is the fluorescence intensity in the absence of RNA, a is proportional to the ratio of the quantum yields of the 2AP-RNA complex and free 2AP, and K_{Dapp} is the apparent dissociation constant. The quantum yield of free 2AP is higher than the yield for the complex and thus, $a < 1$.

3.5 Figures

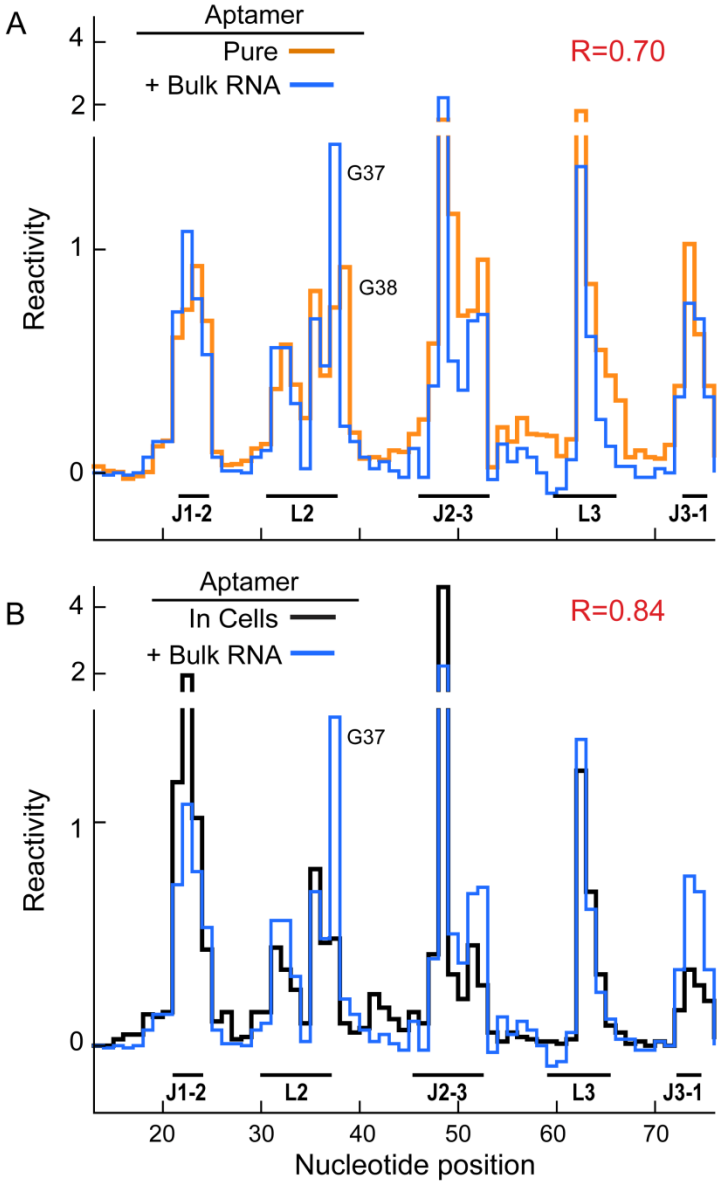

Figure 3.1 SHAPE profiles of the aptamer in the presence of total cellular RNA (400 ng/uL and 1 mM Mg^{2+}) compared to those obtained in (A) dilute buffer with 1 mM Mg^{2+}, and (B) in cells. For in vitro experiments the aptamer concentration was 1 μM.

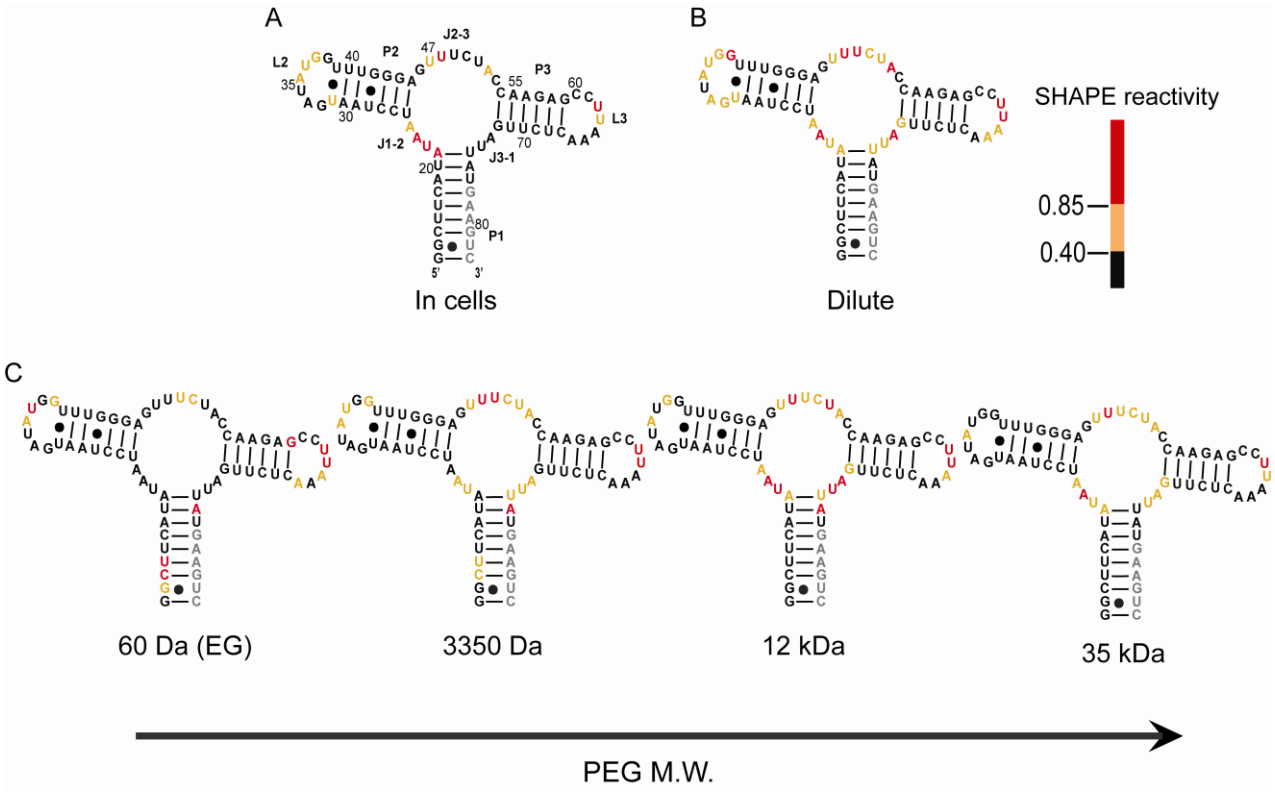

Figure 3.2 SHAPE reactivities for the free aptamer in (A) cells, (B) buffer with1 mM Mg^{2+} and (C) buffer with 1 mM Mg^{2+} and PEG 100 g/L, superimposed on the secondary structure of the aptamer. Highly reactive nucleotides are colored red, moderately reactive, orange, and unreactive, black. Nucleotides colored in grey have no data due to their proximity to the reverse transcription primer binding site.

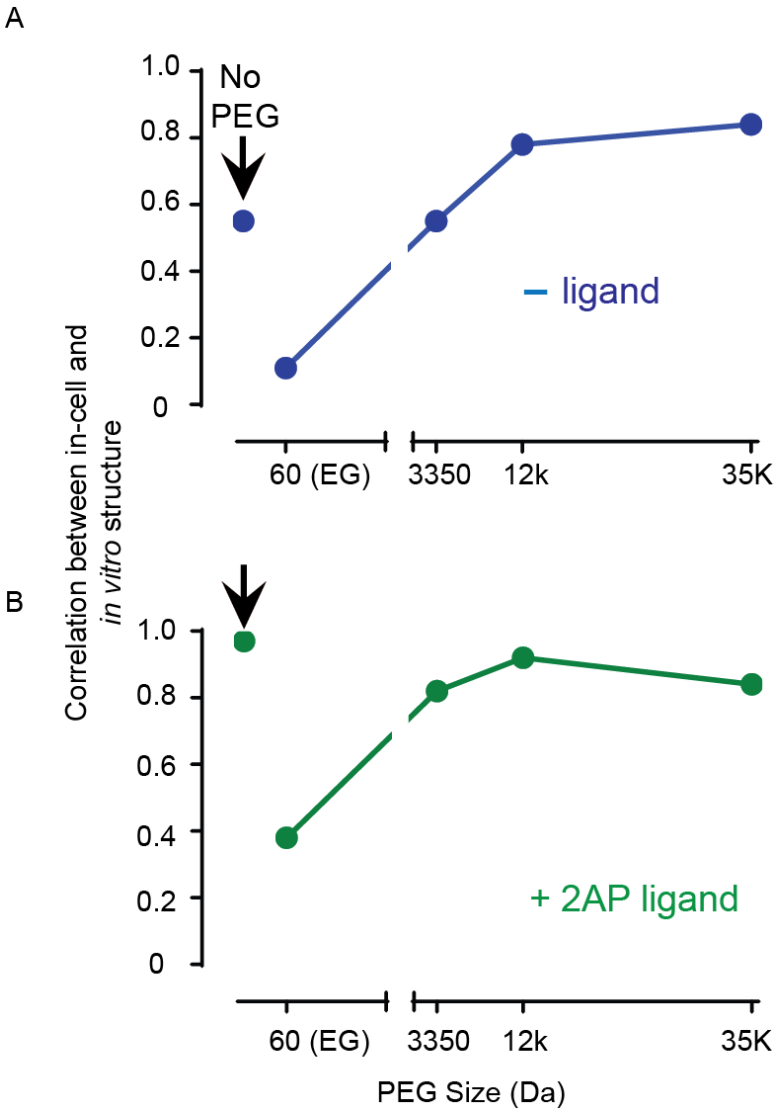

Figure 3.3 Correlation between the in-cell SHAPE profile (Chapter 2) and profiles obtained *in vitro* in the presence of PEG, shown as a function of PEG size for the (A) free and (B) bound aptamer. The correlation value for the in-cell versus dilute buffer (no PEG) is marked by an arrow.

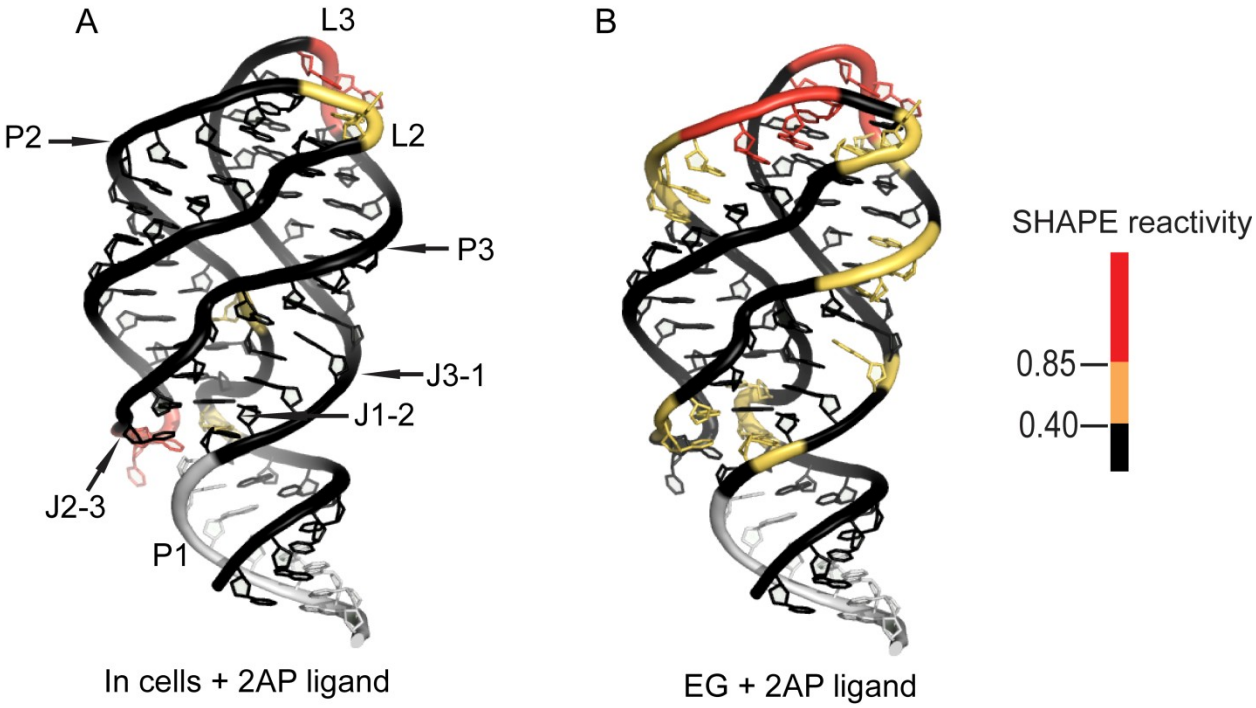

Figure 3.4 SHAPE reactivities for the ligand-bound aptamer in (A) cells, and (B) buffer with EG superimposed on the structure for the adenine-bound aptamer (PDB 1Y26).[13] Highly reactive nucleotides are red, moderately reactive are orange, and unreactive are colored black. Nucleotides colored in grey have no data due to their proximity to the reverse transcription primer binding site.

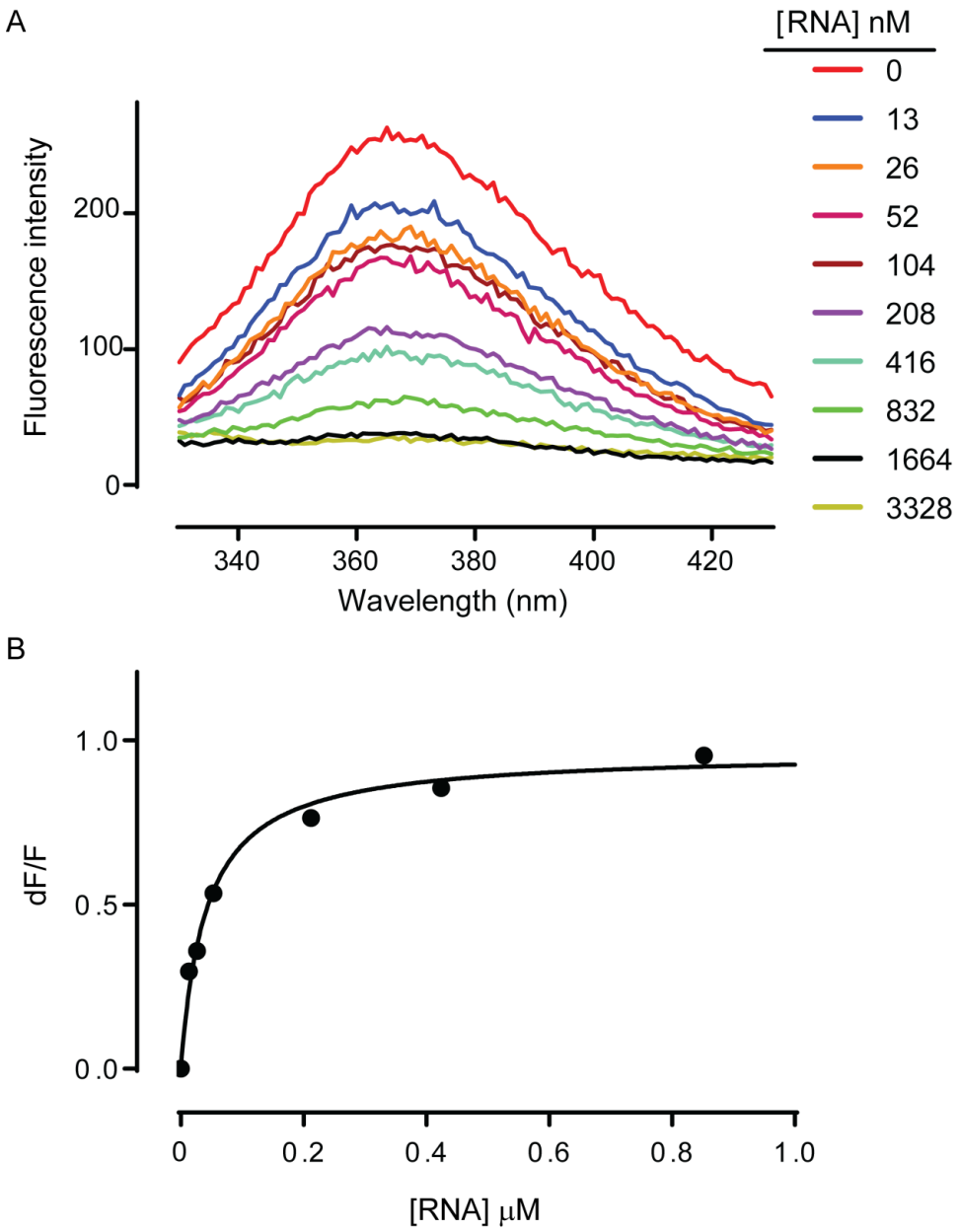

Figure 3.5 Determining $K_{Dapp,\ dilute.}$ (A) Fluorescence quenching of 2AP (40 nM) with aptamer RNA as titrant. (B) Peaks were integrated and fit to a single-site binding function[45] (Equation 3.1) to yield a $K_{Dapp,\ dilute}$ of 48 ± 5 nM.

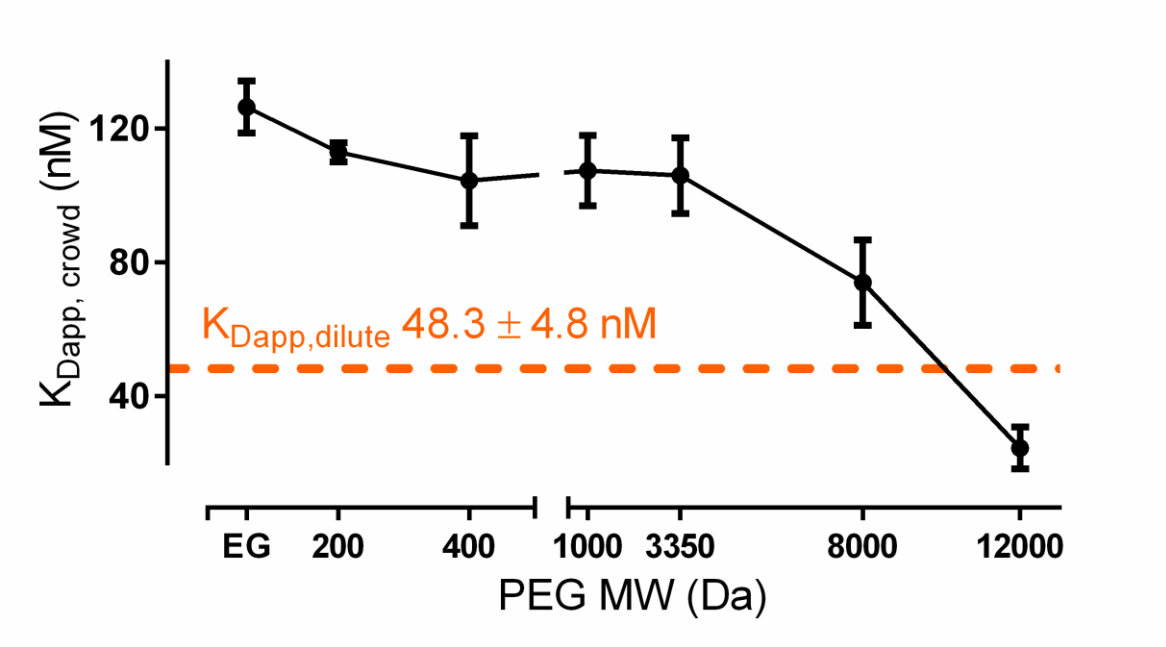

Figure 3.6 $K_{Dapp,\ crowd}$ measured as a function of PEG molecular weight. For comparison, K_{Dapp} measured in buffer alone is shown as a dashed orange line.

Chapter 4: Pitfalls and Future Directions

4.1 Determining K_D in cells using SHAPE

In-cell SHAPE studies of the *add* A aptamer show that the conformation of the aptamer in cells is significantly different than that in dilute buffer *in vitro* (Figure 2.5). Overall, the conformation in cells is more compact and ordered. These structural differences likely result in a reduced free energy requirement for switching from the free to ligand-bound state in cells compared to in buffer. Ultimately, the reported K_D (~117 nM), measured in dilute buffer, may not be physiologically relevant.

Binding studies using in-cell SHAPE revealed that the extent of conformational change for the aptamer is dependent on the concentration of ligand added to cells. Concentration dependent SHAPE data for several individual nucleotides, which undergo a conformational change upon ligand binding, fit well to a one-site binding equation (Figure 4.1). These results suggest that an *in vivo* equilibrium dissociation constant for the aptamer-2AP complex can be measured using SHAPE.

Two caveats arose in using these data to determine the K_D for the aptamer-2AP complex in cells: 1) the concentration of 2-aminopurine is representative only of that added to the media and not of that inside the cells, and 2) to accurately measure K_D using the receptor signal (in this case, the aptamer conformational

change), the concentration of receptor must be << K_D, and held constant. Under our expression conditions in BL21(DE3) *E. coli* cells, the aptamer concentration is estimated to be ~10^{-3} M, much higher than the K_D for the aptamer-2AP complex (~100 nM, measured *in vitro*[46]).

4.1.1 Measuring 2-aminopurine concentration in cells

Development of an accurate method for determination of 2AP in cells will be required for in-cell K_D measurements. Attempts to measure the 2AP concentration in cells relied on detection of 2AP fluorescence inside cells. Two methods were applied: 1) multi-photon excitation fluorescence (MPEF) microscopy, and 2) fluorescence spectroscopy.

MPEF microscopy is superior to single-photon excitation microscopy for imaging of live cells,[52] and has been used previously to investigate 2AP fluorescent properties.[53] However, in my studies, fluorescence signal for pure 2AP, up to molar concentrations, could not be detected above background using either 2PEF or 3PEF microscopy. Based on these results, in-cell measurements were deemed unfeasible and not attempted.

To apply fluorescence spectroscopy, cells were incubated with 2AP at varying concentrations, washed, and fluorescence intensity was measured (λ_{ex}=300 nm, λ_{em} 330 nm – 450 nm) for cell aliquots using a Varian Cary Eclipse fluorimeter. Although results indicated that some level of fluorescence could be detected over background autofluorescence, the signal was low and not consistently dependent on the concentration of 2AP added to cell media. Lysing cells prior to fluorescence measurements gave results similar to those observed with intact cells. To determine

K_D in cells, a method is required to accurately measure the concentration of 2AP in cells.

4.1.1.1 Mass spectrometry to measure 2-aminopurine concentration in cells

Mass spectrometry has been used to measure metabolite concentrations in *E. coli*.[47] Application of this method may be promising for measuring 2AP concentrations in cells. To construct a binding curve and determine K_D in cells, at least 7 cellular 2AP concentrations will be required, where the concentration of 2AP equal to K_D falls near the midpoint of titration. This can be achieved by adjusting the concentration of 2AP added to cell media and optimizing incubation times.

4.1.2 Controlling aptamer expression for in-cell K_D measurements

In our system, expression of the aptamer-tRNA construct is controlled by the *E. coli lac* operon repressor. Gene expression is induced by the addition of the lactose analog isopropyl β-D-1-thiogalactopyranoside (IPTG). The *lac* operon is regulated by a positive feedback loop.[54] Induction of the *lac* operon by IPTG turns on expression of the gene encoding the lactose transport protein, permease, which leads to increased uptake of IPTG. Consequently, inducing expression with less-than-saturating concentrations of IPTG can lead to a heterogeneous culture, where some population of cells has expression turned ‘on’ and others turned ‘off’.

Accurate in-cell binding studies using SHAPE require homogenous expression at levels much lower than can be achieved with addition of saturating IPTG. To overcome the consequences of the *lac* operon positive feedback loop, and fulfill the requirements for binding studies, I used Tuner™(DE3) *E. coli* cells (Novagen®). These cells are mutants of BL21 strains that contain a *lac* permease

(*lacY*) mutation, which allows free diffusion of IPTG into the entire population of cells, thus allowing concentration dependent and homogeneous induction levels. Using Tuner™(DE3), I have shown that the expression level of the aptamer construct is linear with respect to IPTG concentration and can be adjusted from very low levels to robust levels, as seen with saturating IPTG in BL21(DE3) *E. coli* cells. To accurately measure KD in cells aptamer expression levels must be quantified and optimized.

4.1.2.1 Quantifying and optimizing low-level (≤ 10^{-9} M) aptamer expression

Expression levels of the *add* A aptamer-tRNA construct required for binding are too low to be visualized by gel electrophoresis. Real-time PCR is a standard method for measuring gene expression and has a detection limit of a mere three template copies.[55] To determine the concentration of the aptamer construct in cells, real-time PCR would be applied to total cellular RNA harvested following induction with varying concentrations of IPTG. Comparison to real-time PCR results from a standard curve of purified aptamer construct with known concentrations, the concentration of aptamer construct in total cellular RNA can be determined. The concentration in cells can be calculated from the following equation:

$$\frac{RNA \ \times V_{total\ RNA}}{\#\ cells\ \times V_{cells}} = \frac{RNA\ (mol)}{cytosol\ (L)} \qquad (2)$$

Where V is the volume of the total RNA or of cells and the # cells can be determined from the optical density at 600 nm, measured immediately before RNA harvest.[56] For in-cell binding studies, expression time and IPTG concentration should be optimized such that $(RNA\ (mol))/(cytosol\ (L))$ << K_D. Precise timing control can be achieved by adding rifampicin to cells, which halts transcription.[57]

4.2 Production of ^{19}F-labeled RNA for *in vitro* or in-cell NMR

One-dimensional ^{19}F nuclear magnetic resonance spectroscopy (NMR) has proved to be useful in studying RNA secondary structure.[58] Primary methods to obtain ^{19}F labeled RNA include chemical or enzymatic synthesis. Chemical synthesis is costly and is only practical for RNAs < 55 nt.[59] Enzymatic synthesis, using T7 polymerase, can be applied to produce larger RNAs, however, NMR requires milligram quantities of labeled RNA. To produce such quantities, atypically large transcription reactions (> 50 ml per NMR experiment) containing mM concentrations of nucleotide triphosphates (NTPs), are required. In my experience; ^{19}F-labeled nucleotides are not readily available in large quantities and are expensive, making enzymatic synthesis of ^{19}F-labeled RNA unfeasible. It would be useful to have a high-yield, cost efficient method for producing monodisperse ^{19}F-labeled RNA.

The tRNA scaffold expression system, described in Chapter 1, may present a unique opportunity to solve the ^{19}F-labeling problem, as it is a reliable method for producing large quantities of ^{15}N or ^{13}C-labeled RNA simply by using isotope labeled media.[30,31] It has been shown that *E. coli* cells will take up 5-fluorouracil (5FU), convert it to 5-fluorouridine 5'-triphosphate (5'-fluoro-UTP), and incorporate it in place of uracil in RNA transcription.[60] In addition to substitution of RNA uracil nucleotides, a second consequence of 5FU exposure to cells is inhibition of thymidylate synthetase, which is responsible for the conversion of deoxyuridine monophosphate to deoxythymidine monophosphate.[61] The resultant depletion of deoxythymidine triphosphate disrupts DNA metabolism, and halts cell growth. In LB

media, substituted with 5FU, cell growth can be recovered by addition of thymidine at an equal concentration to 5FU.[62]

NMR spectra from $tRNA^{val}$, isolated from *E. coli* grown in glucose-salts medium in the presence of 5FU and thymimdine, show > 90% substitution of uracil nucleotides.[63,64] These observations suggest that RNA overexpression, via the tRNA scaffold system, in the presence of 5FU could yield large quantities of ^{19}F labeled RNA, suitable for NMR analysis. In addition, previous studies using a mere 100 MHz NMR spectrometer, indicate that 5FU substituted RNA can be detected in *E. coli* cells.[65]

4.2.1 ^{19}F labeled aptamer construct cannot be detected by in-cell NMR

E. coli cells were grown in M9 minimal media. When the culture $O.D._{600nm}$ reached ~0.6, expression of the aptamer construct was induced by addition of IPTG, and 5FU (25 mg/ml) and thymidine (25 mg/ml) were added. Expression continued for 3 hours and cells were centrifuged, washed 3X, and resuspended in 2 ml of media. Lysates were prepared from sonicated resuspended cell pellets and supernatants were collected after centrifugation of lysates. ^{19}F spectra of cells, lysates, and supernatants were acquired at 37° C on a Varian Inova 600 MHz spectrometer equipped with a 5 mm $^{19}F(H)$ z-gradient probe.

NMR of cells and lysate showed that 5FU substituted aptamer/tRNA construct could not be detected using ^{19}F NMR (Fig. 4.2). Free 5FU and a metabolic product of 5FU (likely 5-fluorouridine based on the ~5 ppm shift[65]) could be detected in cells, lysate, and supernatant at varying concentrations. The absence of peaks representing the overexpressed aptamer construct is likely due to the large size (143

nt) and therefore slow tumbling of the aptamer construct, particularly in the high viscosity lysate or cellular environment. The tRNA scaffold expression system is not suitable for in cell NMR studies.

4.2.2 *In vitro* ^{19}F NMR studies: low yield expression and increased susceptibility to degradation

The aptamer construct can purified from total cellular RNA using anion exchange FPLC (Fig. 4.3A and 4.3B). Following purification, the tRNA scaffold can be removed by annealing DNA oligonucleotides and subsequent RNaseH digestion to yield a monodisperse product (Fig 4.3C).[30,31] Removing the tRNA scaffold not only reduces the number of possible resonances, but allows specific analysis of the aptamer domain.

Two problems arose in development of this method. First, due to slow cell growth in M9 minimal media substituted with 5FU, more than 12 L of culture is required to yield enough aptamer construct to be detected by ^{19}F NMR, making this method time-inefficient. To efficiently produce the large amounts of ^{19}F-labeled RNA required for NMR (>10 mg), aptamer expression must be optimized such that yields are significantly higher than previously observed. This could be achieved by adjusting expression time and media formula.

Second, following RNaseH digestion, purification of the aptamer from excess DNA oligonucleotides and RNA fragments by FPLC, resulted in total loss of aptamer due to degradation. Degradation of the aptamer was likely due to RNase contamination in the FPLC. To avoid aptamer degradation, an FPLC dedicated to RNA purification would be required.

4.3 Time-resolved SHAPE to measure crowding effects on RNA folding

Upon ligand binding, the *add* A aptamer folds into a highly organized structure in discreet steps over a ~ 2 minute time period.[16] These steps include docking of the ligand in the binding pocket, docking of L2 and L3 and organization of the binding pocket, followed by tightening of the P1 stem, and finally, full stabilization of the L2-L3 tertiary interactions and binding pocket to achieve the final bound structure. Results from in-cell structural studies (Chapter 2) suggest that crowding may perturb the folding pathway demonstrated in dilute solution. First, the tertiary contacts for L2 and L3 are in place under crowded conditions in cells and undergo little change upon ligand binding, suggesting they are fully formed. Second, the binding pocket exhibits a higher level of organization in cells compared to in buffer. These observations suggest that the folding pathway for the aptamer will be altered under crowded conditions.

Time-resolved SHAPE is an ideal method to monitor the time-dependent conformational change for the aptamer. This method employs the SHAPE reagent benzoyl cyanide (BzCN), which reacts completely with RNA in ~1 s at 37° C.[66] Therefore, BzCN can yield information on RNA folding on a seconds time scale. The experiment would involve initiating folding by addition of ligand to the aptamer RNA followed by subsequent additions of BzCN, and removal of aliquots 1 s after each addition until RNA folding is presumably complete (~ 3 min). Time resolution of aptamer folding will be limited to the efficiency of the user. Aliquots from each time-point are analyzed as described for *in vitro* SHAPE in Chapter 2.

Analysis of the aptamer ligand-induced folding pathway under crowded conditions can be achieved by carrying out time-resolved SHAPE in the presence of synthetic crowding agents with various concentrations, such as those described in Chapter 3. These results will improve our understanding of the effects of crowding on the folding kinetics and structural transitions for *add* A aptamer ligand binding.

4.4 Figures

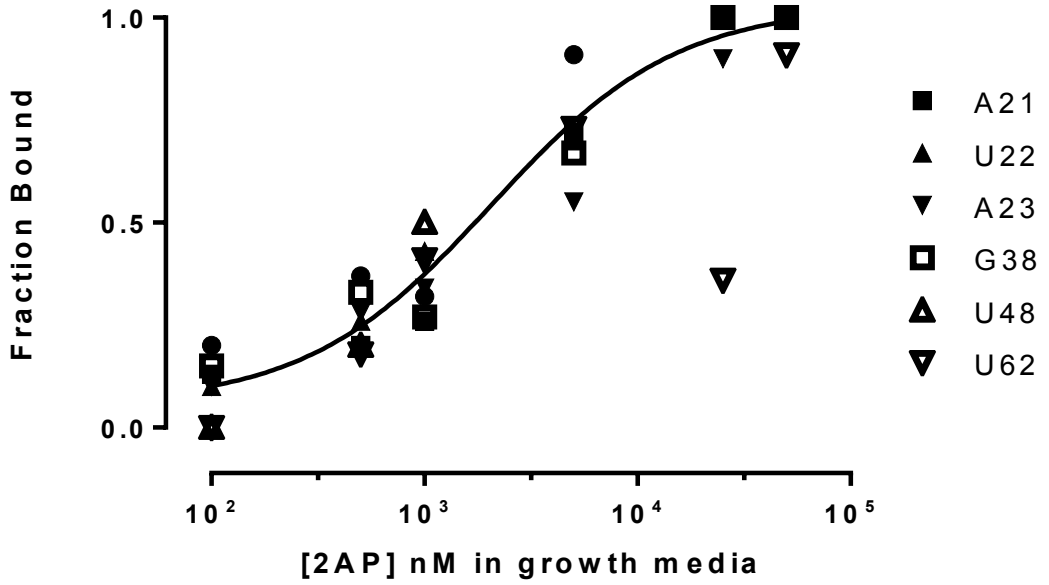

Figure 4.1 Ligand binding can be detected in cells using SHAPE. 2AP was added to cell growth media at varying concentrations. Concentration dependent conformational changes can be detected for several nucleotides in the *add* A aptamer domain using SHAPE. These changes fit well to a one site binding curve.

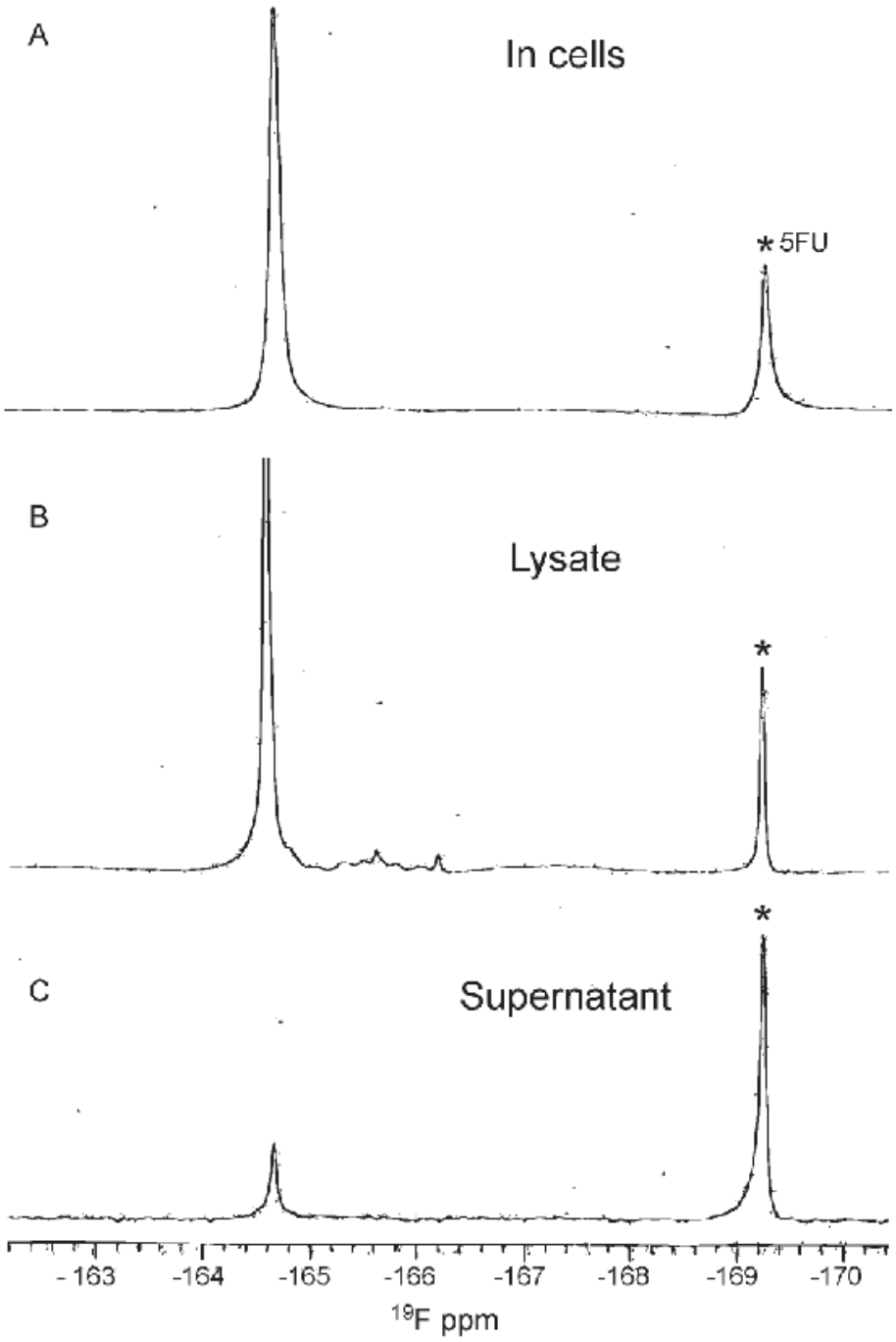

Figure 4.2 Aptamer construct cannot be detected in cells or in lysate using ^{19}F NMR. ^{19}F spectra of (A) E. coli cells overexpressing the aptamer construct grown in 5FU supplemented media, (B) lysates, and (C) supernatant. Asterisks are free 5FU resonances.

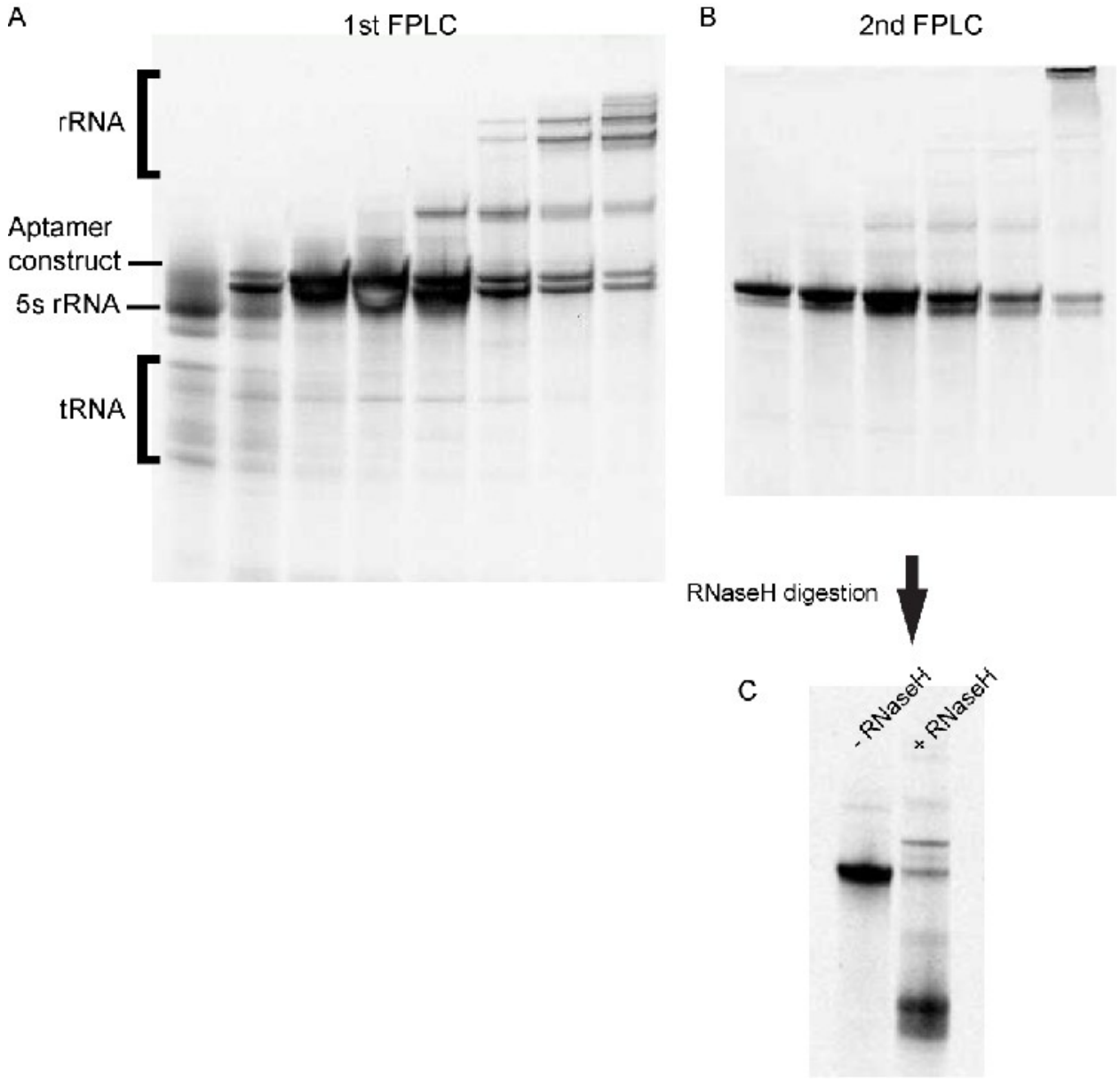

Figure 4.3 FPLC purification and RNaseH digestion of the aptamer construct. (A) Total cellular RNA was isolated and subjected to anion exchange FPLC. (B) Fractions containing the aptamer construct were run a second time on FPLC for further purification. (C) DNA oligonucleotides, complementary to the tRNA scaffold, were annealed to the pure construct, which was then digested with RNaseH to yield the *add* A aptamer domain.

References

1. Zimmerman, S. B., and Trach, S. O. (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli, *J. Mol. Biol. 222*, 599-620.

2. Minton, A. P. (1981) Excluded volume as a determinant of macromolecular structure and reactivity, *Biopolymers 20*, 2093-2120.

3. Zhou, H. X., Rivas, G., and Minton, A. P. (2008) Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences, *Annu. Rev. Biophys. 37*, 375-397.

4. Minton, A. P., and Wilf, J. (1981) Effect of macromolecular crowding upon the structure and function of an enzyme: glyceraldehyde-3-phosphate dehydrogenase, *Biochemistry 20*, 4821-4826.

5. Benton, L. A., Smith, A. E., Young, G. B., and Pielak, G. J. (2012) Unexpected effects of macromolecular crowding on protein stability, *Biochemistry 51*, 9773-9775.

6. Denesyuk, N. A., and Thirumalai, D. (2011) Crowding promotes the switch from hairpin to pseudoknot conformation in human telomerase RNA, *JACS 133*, 11858-11861.

7. Kilburn, D., Roh, J. H., Guo, L., Briber, R. M., and Woodson, S. A. (2011) Molecular crowding stabilizes folded RNA structure by the excluded volume effect, *JACS 132*, 8690-8696.

8. Hall, K. B. (2008) RNA in motion, *Curr. Opin. Chem. Biol. 12*, 612-618.

9. Pincus, D. L., Hyeon, C., and Thirumalai, D. (2008) Effects of trimethylamine N-oxide (TMAO) and crowding agents on the stability of RNA hairpins, *JACS 130*, 7364-7372.

10. Nakano, S., Karimata, H. T., Kitagawa, Y., and Sugimoto, N. (2009) Facilitation of RNA enzyme activity in the molecular crowding media of cosolutes, *JACS 131*, 16881-16888.

11. Mandal, M., and Breaker, R. R. (2004) Gene regulation by riboswitches, *Nat. Rev. Mol. Cell. Biol. 5*, 451-463.

12. Mandal, M., and Breaker, R. R. (2004) Adenine riboswitches and gene activation by disruption of a transcription terminator, *Nat. Struct. Mol. Biol. 11*, 29-35.

13. Serganov, A., Yuan, Y. R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A. T., Hobartner, C., Micura, R., Breaker, R. R., and Patel, D. J. (2004) Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs, *Chem. Biol. 11*, 1729-1741.

14. Serganov, A., and Patel, D. J. (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins, *Nat. Rev. Genet. 8*, 776-790.

15. Dixon, N., Duncan, J. N., Geerlings, T., Dunstan, M. S., McCarthy, J. E., Leys, D., and Micklefield, J. (2010) Reengineering orthogonally selective riboswitches, *Proc. Natl. Acad. Sci. USA 107*, 2830-2835.

16. Lee, M. K., Gal, M., Frydman, L., and Varani, G. (2010) Real-time multidimensional NMR follows RNA folding with second resolution, *Proc. Natl. Acad. Sci. USA 107*, 9192-9197.

17. Wells, S. E., Hughes, J. M., Igel, A. H., and Ares, M., Jr. (2000) Use of dimethyl sulfate to probe RNA structure in vivo, *Methods Enzymol 318*, 479-493.

18. Liebeg, A., and Waldsich, C. (2009) Probing RNA structure within living cells, *Methods Enzymol 468*, 219-238.

19. Merino, E. J., Wilkinson, K. A., Coughlan, J. L., and Weeks, K. M. (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE), *JACS 127*, 4223-4231.

20. Wang, B., Wilkinson, K. A., and Weeks, K. M. (2008) Complex ligand-induced conformational changes in tRNA(Asp) revealed by single-nucleotide resolution SHAPE chemistry, *Biochemistry 47*, 3454-3461.

21. Weeks, K. M., and Mauger, D. M. (2011) Exploring RNA structural codes with SHAPE chemistry, *Acc. Chem. Res. 44*, 1280-1291.

22. Grohman, J. K., Gorelick, R. J., Lickwar, C. R., Lieb, J. D., Bower, B. D., Znosko, B. M., and Weeks, K. M. (2013) A guanosine-centric mechanism for RNA chaperone function, *Science 340*, 190-195.

23. Hajdin, C. E., Bellaousov, S., Huggins, W., Leonard, C. W., Mathews, D. H., and Weeks, K. M. (2013) Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots, *Proc. Natl. Acad. Sci. U S A 110*, 5498-5503.

24. Spitale, R. C., Crisalli, P., Flynn, R. A., Torre, E. A., Kool, E. T., and Chang, H. Y. (2013) RNA SHAPE analysis in living cells, *Nat. Chem. Biol. 9*, 18-20.

25. Steen, K. A., Rice, G. M., and Weeks, K. M. (2012) Fingerprinting noncanonical and tertiary RNA structures by differential SHAPE reactivity, *JACS 134*, 13160-13163.

26. Leipply, D., and Draper, D. E. (2011) Effects of Mg^{2+} on the free energy landscape for folding a purine riboswitch RNA, *Biochemistry 50*, 2790-2799.

27. Sarkar M., L. C., Pielak G.J. (2013) Soft interactions and crowding, *Biophys. Rev. 5*, 187-194.

28. Wang, Y., Sarkar, M., Smith, A. E., Krois, A. S., and Pielak, G. J. (2012) Macromolecular crowding and protein stability, *JACS 134*, 16614-16618.

29. Draper, D. E. (2004) A guide to ions and RNA structure, *RNA 10*, 335-343.

30. Ponchon, L., and Dardel, F. (2007) Recombinant RNA technology: the tRNA scaffold, *Nat. Methods 4*, 571-576.

31. Ponchon, L., Beauvais, G., Nonin-Lecomte, S., and Dardel, F. (2009) A generic protocol for the expression and purification of recombinant RNA in Escherichia coli using a tRNA scaffold, *Nat. Protoc. 4*, 947-959.

32. Deigan, K. E., Li, T. W., Mathews, D. H., and Weeks, K. M. (2009) Accurate SHAPE-directed RNA structure determination, *Proc. Natl. Acad. Sci. USA 106*, 97-102.

33. Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess, J. W., Jr., Swanstrom, R., Burch, C. L., and Weeks, K. M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome, *Nature 460*, 711-716.

34. Rieder, R., Lang, K., Graber, D., and Micura, R. (2007) Ligand-induced folding of the adenosine deaminase A-riboswitch and implications on riboswitch translational control, *ChemBioChem 8*, 896-902.

35. Mortimer, S. A., and Weeks, K. M. (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry, *JACS 129*, 4144-4145.

36. Grynkiewicz, G., Poenie, M., and Tsien, R. Y. (1985) A new generation of Ca^{2+} indicators with greatly improved fluorescence properties, *J. Biol. Chem. 260*, 3440-3450.

37. Sechan Wee, B. J. W. (1988) Insights into the cell envelope of *Paracoccus denitrificans*, a member of the α-subdivision of purple bacteria, through studies of its lysozyme susceptibility, *Can. J. Microbiol. 34*, 952-959.

38. London, R. E. (1991) Methods for measurement of intracellular magnesium: NMR and fluorescence, *Annu. Rev. Physiol. 53*, 241-258.

39. Hurley, T. W., Ryan, M. P., and Brinck, R. W. (1992) Changes of cytosolic Ca^{2+} interfere with measurements of cytosolic Mg^{2+} using mag-fura-2, *Am. J. Physiol. 263*, C300-307.

40. Tashiro, M., and Konishi, M. (1997) Basal intracellular free Mg^{2+} concentration in smooth muscle cells of guinea pig tenia cecum: intracellular calibration of the fluorescent indicator furaptra, *Biophys. J. 73*, 3358-3370.

41. Froschauer, E. M., Kolisek, M., Dieterich, F., Schweigel, M., and Schweyen, R. J. (2004) Fluorescence measurements of free [Mg^{2+}] by use of mag-fura 2 in Salmonella enterica, *FEMS Microbiol. Lett. 237*, 49-55.

42. Gherghe, C. M., Mortimer, S. A., Krahn, J. M., Thompson, N. L., and Weeks, K. M. (2008) Slow conformational dynamics at C2'-endo nucleotides in RNA, *JACS 130*, 8884-8885.

43. Mortimer, S. A., and Weeks, K. M. (2009) C2'-endo nucleotides as molecular timers suggested by the folding of an RNA domain, *Proc. Natl. Acad. Sci. U S A 106*, 15622-15627.

44. McGinnis, J. L., Dunkle, J. A., Cate, J. H., and Weeks, K. M. (2012) The mechanisms of RNA SHAPE chemistry, *JACS 134*, 6617-6624.

45. Lemay, J. F., Penedo, J. C., Tremblay, R., Lilley, D. M., and Lafontaine, D. A. (2006) Folding of the adenine riboswitch, *Chem. Biol. 13*, 857-868.

46. Lemay, J. F., and Lafontaine, D. A. (2007) Core requirements of the adenine riboswitch aptamer for ligand binding, *RNA 13*, 339-350.

47. Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., and Rabinowitz, J. D. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli, *Nat. Chem. Biol. 5*, 593-599.

48. Gherghe, C. M., Shajani, Z., Wilkinson, K. A., Varani, G., and Weeks, K. M. (2008) Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S2) in RNA, *JACS 130*, 12244-12245.

49. Cayley, S., Lewis, B. A., Guttman, H. J., and Record, M. T., Jr. (1991) Characterization of the cytoplasm of Escherichia coli K-12 as a function of external osmolarity. Implications for protein-DNA interactions in vivo, *J Mol Biol 222*, 281-300.

50. Karabiber, F., McGinnis, J. L., Favorov, O. V., and Weeks, K. M. (2013) QuShape: Rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis, *RNA 19*, 63-73.

51. Knowles, D. B., LaCroix, A. S., Deines, N. F., Shkel, I., and Record, M. T., Jr. (2011) Separation of preferential interaction and excluded volume effects on DNA duplex and hairpin stability, *Proc. Natl. Acad. Sci. U S A 108*, 12699-12704.

52. Oheim, M., Michael, D. J., Geisbauer, M., Madsen, D., and Chow, R. H. (2006) Principles of two-photon excitation fluorescence microscopy and other nonlinear imaging approaches, *Adv. Drug Deliv. Rev. 58*, 788-808.

53. Katilius, E., and Woodbury, N. W. (2006) Multiphoton excitation of fluorescent DNA base analogs, *J. Biomed. Opt. 11*, 044004.

54. van Hoek, M., and Hogeweg, P. (2007) The effect of stochasticity on the lac operon: an evolutionary perspective, *PLoS Comput. Biol. 3*, e111.

55. Bernardo, V., Ribeiro Pinto, L. F., and Albano, R. M. (2013) Gene expression analysis by real-time PCR: experimental demonstration of PCR detection limits, *Anal. Biochem. 432*, 131-133.

56. Volkmer, B., and Heinemann, M. (2011) Condition-dependent cell volume and concentration of Escherichia coli to facilitate data conversion for systems biology modeling, *PLoS One 6*, e23126.

57. Sensi, P., Greco, A. M., and Ballotta, R. (1959) Rifomycin. I. Isolation and properties of rifomycin B and rifomycin complex, *Antibiot. Annu. 7*, 262-270.

58. Graber, D., Moroder, H., and Micura, R. (2008) 19F NMR spectroscopy for the analysis of RNA secondary structure populations, *JACS 130*, 17230-17231.

59. Lu, K., Miyazaki, Y., and Summers, M. F. (2010) Isotope labeling strategies for NMR studies of RNA, *J. Biomol. NMR 46*, 113-125.

60. Warner, H. R., and Rockstroh, P. A. (1980) Incorporation and excision of 5-fluorouracil from deoxyribonucleic acid in Escherichia coli, *J. Bacteriol. 141*, 680-686.

61. Oda, Y. (1987) Induction of SOS responses in Escherichia coli by 5-fluorouracil, *Mutat. Res./DNA Rep. Rep. 183*, 103-108.

62. Horowitz, J., Ou, C. N., and Ishaq, M. (1974) Isolation and partial characterization of Escherichia coli valine transfer RNA with uridine-derived residues replaced by 5-fluorouridine, *J. Mol. Biol. 88*, 301-312.

63. Lowrie, R. J., and Bergquist, P. L. (1968) Transfer ribonucleic acids from Escherichia coli treated with 5-fluorouracil, *Biochemistry 7*, 1761-1770.

64. Hardin, C. C., Gollnick, P., Kallenbach, N. R., Cohn, M., and Horowitz, J. (1986) Fluorine-19 nuclear magnetic resonance studies of the structure of 5-fluorouracil-substituted Escherichia coli transfer RNA, *Biochemistry 25*, 5699-5709.

65. Gochin, M., James, T. L., and Shafer, R. H. (1984) In vivo 19F-NMR of 5-fluorouracil incorporation into RNA and metabolites in Escherichia coli cells, *BBA-Mol. Cell Res. 804*, 118-124.

66. Mortimer, S. A., and Weeks, K. M. (2008) Time-resolved RNA SHAPE chemistry, *JACS 130*, 16178-16180.